TAB J

beled at the 5′ end were purified by Sephadex G-50 chromatography in a 0.3*M* sodium acetate, 0.01*M* tris-HCl, *p*H 7.9, 0.001*M* EDTA system, and concentrated by precipitation with ethanol.
6. A. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
7. Low molecular weight DNA fragments were resolved by electrophoresis at < 10 volt/cm in gels containing 10 percent polyacrylamide (acrylamide: *N,N′*-methylenebisacrylamide = 25 : 1) in one-half strength TBE [0.05*M* tris-borate, *p*H 8.3, 0.001*M* EDTA, A. Peacock and C. Dingman, *Biochemistry* **7**, 668 (1968)] with 25 percent glycerol. Glycerol improves band sharpness dramatically.
8. J. Dahlberg and F. Blattner, in *Virus Research*, C. Fox and W. Robinson, Eds. (Academic Press, New York, 1973), p. 533.
9. M. Rosenberg, personal communication.
10. D. Kleid, Z. Humayun, A. Jeffrey, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 293 (1976).
11. M. Rosenberg, B. de Crombrugghe, R. Musso, *ibid.*, p. 717.
12. G. Scherer, G. Hobom, H. Kössel, *Nature (London)* **256**, 117 (1977).
13. V. Pirotta, personal communication.
14. E. Schwarz, G. Scherer, G. Hobom, H. Kössel, personal communication.
14a. G. Scherer, G. Hobom, H. Kössel, personal communication.
15. B. Westmoreland, W. Szybalski, H. Ris, *Science* **163**, 1343 (1969).
16. D. Hewish, K. Smith, J. Spencer, *Virology* **74**, 363 (1976).
17. A. Rambach, *ibid.* **54**, 270 (1973).
18. W. Dove, H. Inokuchi, W. Stevens, in *The Bacteriophage Lambda*, A. D. Hershey, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1971), p. 747.
19. W. Szybalski *et al.*, *J. Cell. Physiol.* **74** (Suppl. 1), 33 (1969).
20. H. Inokuchi, W. Dove, D. Freifelder, *J. Mol. Biol.* **74**, 721 (1973).
21. N. Kleckner, *Virology* **79**, 174 (1977).
22. M. Schnös and R. Inman, *J. Mol. Biol.* **55**, 31 (1971); A. Kozinski and B. Lindqvist, personal communication.
23. The λ *ori* region, defined by the *ori*⁻ mutations, has two dPy · dPu stretches (18 bp and 18 of 21 bp) in which the pyrimidines are on the *r* strand. The origin of *Col* E1 replication, defined by the site at which deoxyribonucleotides are added onto an RNA primer in vitro, is 20 bp away from a dPy · dPu stretch (21 of 24 bp) with purines on the transcribed strand [J.-I. Tomizawa, H. Ohmori, R. Bird, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 1865 (1976)]. The origin region for phage fd DNA replication, defined by an in vitro RNA primer, contains a dPy · dPu stretch (29 of 32 bp) with pyrimidines on the transcribed strand (C. Gray, R. Sommer, C. Polke, E. Beck, H. Schaller, personal communication). The pyrimidine tract CTC_5 lies in the region of the *A* gene dependent gap in φX174 at the start of RF to RF replication [S. Eisenberg, B. Harbers, C. Hours, D. Denhardt, *J. Mol. Biol.* **99**, 197 (1975)].
24. F. Sanger, G. Air, B. Barrell, N. Brown, A. Coulson, J. Fiddes, C. Hutchison, P. Slocombe, M. Smith, *Nature (London)* **265**, 687 (1977); P. Baas, H. Jansz, R. Sinsheimer, *J. Mol. Biol.* **102**, 633 (1976); S. Eisenberg, B. Harbers, C. Hours, D. Denhardt, *ibid.* **99**, 197 (1975).
25. R. Fisher, *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
26. We believe that the deletions observed in the nucleotide sequences for *r*93, *r*99, and *r*96 correspond to the *ori*⁻ lesions. The *ori*⁻ mutants were isolated without mutagenesis and occur with frequencies inconsistent with multiple mutations. The deletions could not have arisen during cloning, because they are present in the *ori*⁻ phages λ*Nam*7*Nam*53*c*I857*r*93 and λ*Nam*7*Nam*53*c*I857*r*99. The O protein made from the DNA of these phages in vitro is smaller than that made from DNA of the *ori*⁺ parent phage when measured by SDS polyacrylamide gel electrophoresis [J. Yates and M. Furth, unpublished results, using techniques described in J. Yates, W. Gette, M. Furth, M. Nomura, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 689 (1977)]. The size of each deletion is consistent with the changes in molecular weights of the O protein made in vitro and of DNA restriction fragments. Finally, the recombinational map of the *ori*⁻ mutations is consistent with the order of the sequenced deletions (*r*93-*r*99-*r*96), indicating that the deletions are responsible for the *ori*⁻ pheonotype of these mutants (*2*).
27. J. Szpirer, *Mol. Gen. Genet.* **114**, 297 (1972); M. Furth, C. McLeester, W. Dove, in preparation; L. Lambert, R. Thomas, R. Monnat, W. Szybalski, personal communication.
28. By the criterion of heteroduplex pairing, the region of DNA encoding the COOH-terminal half of the O protein appears to be largely conserved in all five lamboid coliphages studied [M. Simon, R. Davis, N. Davidson, in *The Bacteriophage Lambda*, A. D. Hershey, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1971), p. 313; M. Fiandt, Z. Hradecna, H. Lozeron, W. Szybalski, *ibid.*, p. 329]. Complementation analysis shows that *P* gene function can be provided to λ by each of the other four lamboid phages [W. Dove, *Annu. Rev. Genet.* **2**, 305 (1969); J. Szpirer and P. Brachet, *Mol. Gen. Genet.* **108**, 78 (1970)]. In φ80/λ hybrid phages such as *λimm80hy42* the *O* gene (*27*) and O protein iteself have been shown to be hybrid (J. Yates and M. Furth, unpublished observations). There is genetic evidence for direct interaction between O and P proteins [J.-I. Tomizawa, in *The Bacteriophage Lambda*, A. D. Hershey, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1971), p. 549; M. Furth, C. McLeester, W. Dove, in preparation; (*2*)] and between P protein and the *E. coli* replication apparatus [C. Georgopoulos and I. Herskowitz, in *The Bacteriophage Lambda*, A. D. Hershey, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1971), p. 553; H. Saito and H. Uchida, *J. Mol. Biol.* **113**, 1 (1977)].
29. S. Hayes and W. Szybalski, in *DNA Synthesis and Its Regulation*, M. Goulian, P. Hanawalt, C. Fox, Eds. (Benjamin, Menlo Park, Calif., 1975), p. 486.
30. B. Williams, M. Furth, K. Kruger, D. Moore, K. Denniston-Thompson, F. Blattner, in preparation.
30a. W. Dove, E. Hargrove, M. Ohashi, F. Haugli, A. Guha, *Jpn. J. Genet.* **44** (Suppl. 1), 11 (1969).
31. H. Nijkamp, W. Szybalski, M. Ohashi, W. Dove, *Mol. Gen. Genet.* **114**, 80 (1971).
32. This is paper 2172 from the Laboratory of Genetics of the University of Wisconsin and paper No. 7 in the series "Charon phages for DNA Cloning." Paper No. 6 is (*2*). Supported by NIH grant GM21812 to F.R.B., NIH grant CA 07175 to McArdle Laboratory (W. F. Dove), an NSF predoctoral fellowship and an NIH training grant to McArdle Laboratory (M.E.F.), NIH training grant T32 CA09075 (K.D.-T.), and NIH training grant 144-J825 (D.D.M.). We thank W. F. Dove for support and advice, Nigel Godsen for critical reading of the manuscript, Ed Kopetsky and Brenda Dierschke for technical assistance, R. Roberts for hospitality in his laboratory, A. Maxam for providing the detailed sequencing procedures and A. Honigman for discussions concerning AvaX. This work was done under NIH guidelines, which call for EK1, P1.

19 July 1977; revised 20 September 1977

Expression in *Escherichia coli* of a Chemically Synthesized Gene for the Hormone Somatostatin

Abstract. *A gene for somatostatin, a mammalian peptide (14 amino acid residues) hormone, was synthesized by chemical methods. This gene was fused to the* Escherichia coli *β-galactosidase gene on the plasmid pBR322. Transformation of* E. coli *with the chimeric plasmid DNA led to the synthesis of a polypeptide including the sequence of amino acids corresponding to somatostatin. In vitro, active somatostatin was specifically cleaved from the large chimeric protein by treatment with cyanogen bromide. This represents the first synthesis of a functional polypeptide product from a gene of chemically synthesized origin.*

The chemical synthesis of DNA and recombinant DNA methods provide the technology for the design and synthesis of genes that can be fused to plasmid elements for expression in *Escherichia coli* or other bacteria. As a model system we have designed and synthesized a gene for the small polypeptide hormone, somatostatin (Figs. 1 and 2). The major considerations in the choice of this hormone were its small size and known amino acid sequence (*1*), sensitive radioimmune and biological assays (*2*), and its intrinsic biological interest (*3*). Somatostatin is a tetradecapeptide; it was originally discovered in ovine hypothalamic extracts but subsequently was also found in significant quantities in other species and other tissues (*3*). Somatostatin inhibits the secretion of a number of hormones, including growth hormone, insulin, and glucagon. The effect of somatostatin on the secretion of these hormones has attracted attention to its potential therapeutic value in acromegaly, acute pancreatitis, and insulin-dependent diabetes.

The overall construction of the somatostatin gene and plasmid was designed to result in the in vivo synthesis of a precursor form of somatostatin (see Fig. 1). The precursor protein would not be expected to have biological activity, but could be converted to a functional form by cyanogen bromide cleavage (*4*) after cellular extraction. The synthetic somatostatin gene was fused to the lac operon because the controlling sites of this operon are well characterized.

Given the amino acid sequence of somatostatin, one can design from the genetic code a short DNA fragment containing the information for its 14 amino acids (Fig. 2). The degeneracy of the code allows for a large number of possible sequences that could code for the same 14 amino acids. Therefore, the choice of codons was somewhat arbitrary except for the following restrictions. First, amino acid codons known to be favored in *E. coli* for expression of the MS2 genome were used where appropriate (*5*). Second, since the complete sequence would be constructed from a number of overlapping fragments, the fragments were designed to eliminate undesirable inter- and intramolecular pairing. And third, G·C-rich (guanine-cytosine) followed by A·T-rich (adenine-thymine) sequences were avoided since they might terminate transcription (*6*).

Eight oligonucleotides, varying in length from 11 to 16 nucleotides, labeled

in Fig. 2 as A through H, were synthesized by the triester method (7). In addition to the 14 codons for the structural information of somatostatin, several other features were built into the nucleotide sequence. First, to facilitate insertion into plasmid DNA, the 5′ ends have single-stranded cohesive termini for the Eco RI and Bam HI restriction endonucleases. Second, a methionine codon precedes the normal NH_2-terminal amino acid of somatostatin, and the COOH-terminal codon is followed by two nonsense codons.

In the cloning and expression of the synthetic somatostatin gene we used two plasmids. Each plasmid has an Eco RI substrate site at a different region of the β-galactosidase structural gene (see Figs. 3 and 4). The insertion of the synthetic somatostatin DNA fragment into the Eco RI sites of these plasmids brings the expression of the genetic information in that fragment under control of the lac operon controlling elements. After the insertion of the somatostatin fragment into these plasmids, translation should result in a somatostatin polypeptide preceded either by ten amino acids (pSOM1) or by virtually the whole β-ga-

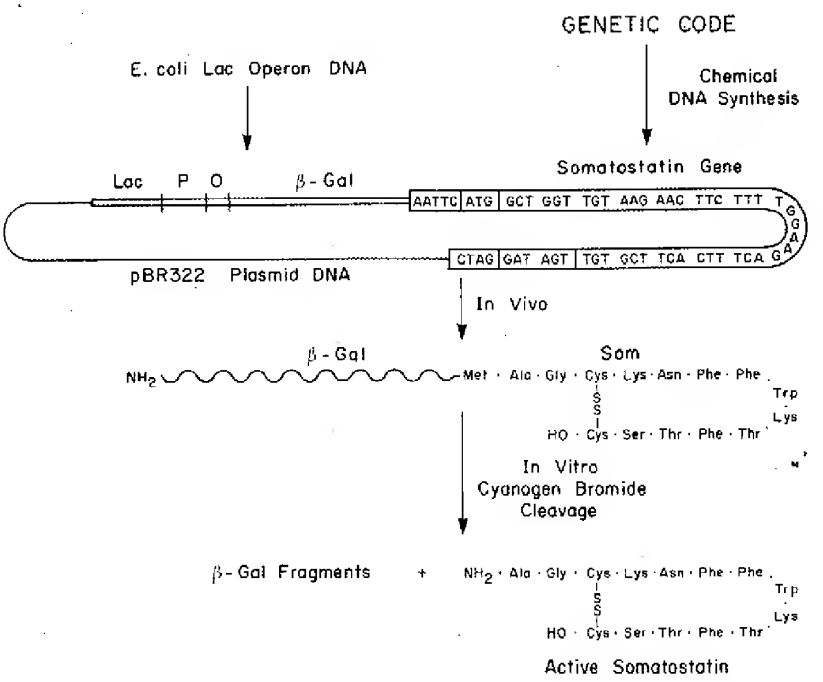
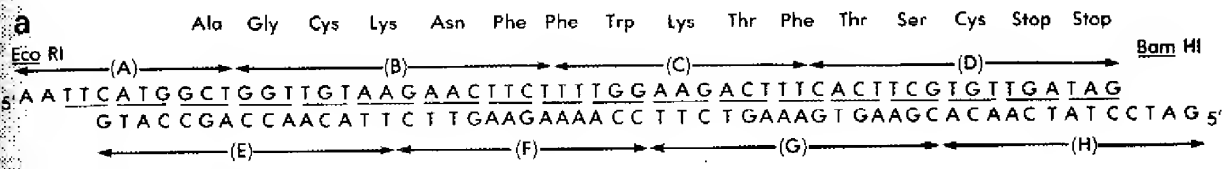


Fig. 1. Schematic outline of the experimental plan. The gene for somatostatin, made by chemical DNA synthesis, was fused to the *E. coli* β-galactosidase gene on the plasmid pBR322. After transformation into *E. coli*, the chimeric plasmid directs the synthesis of a chimeric protein that can be specifically cleaved in vitro at methionine residues by cyanogen bromide to yield active mammalian peptide hormone.



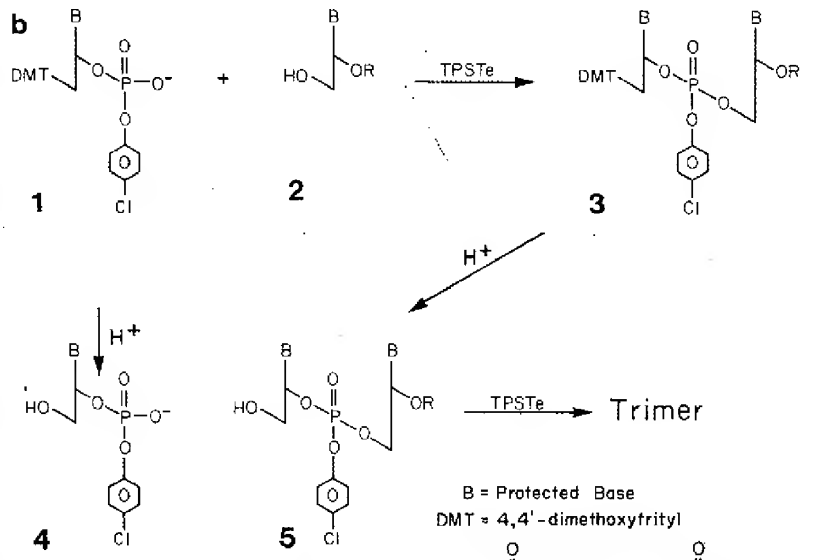
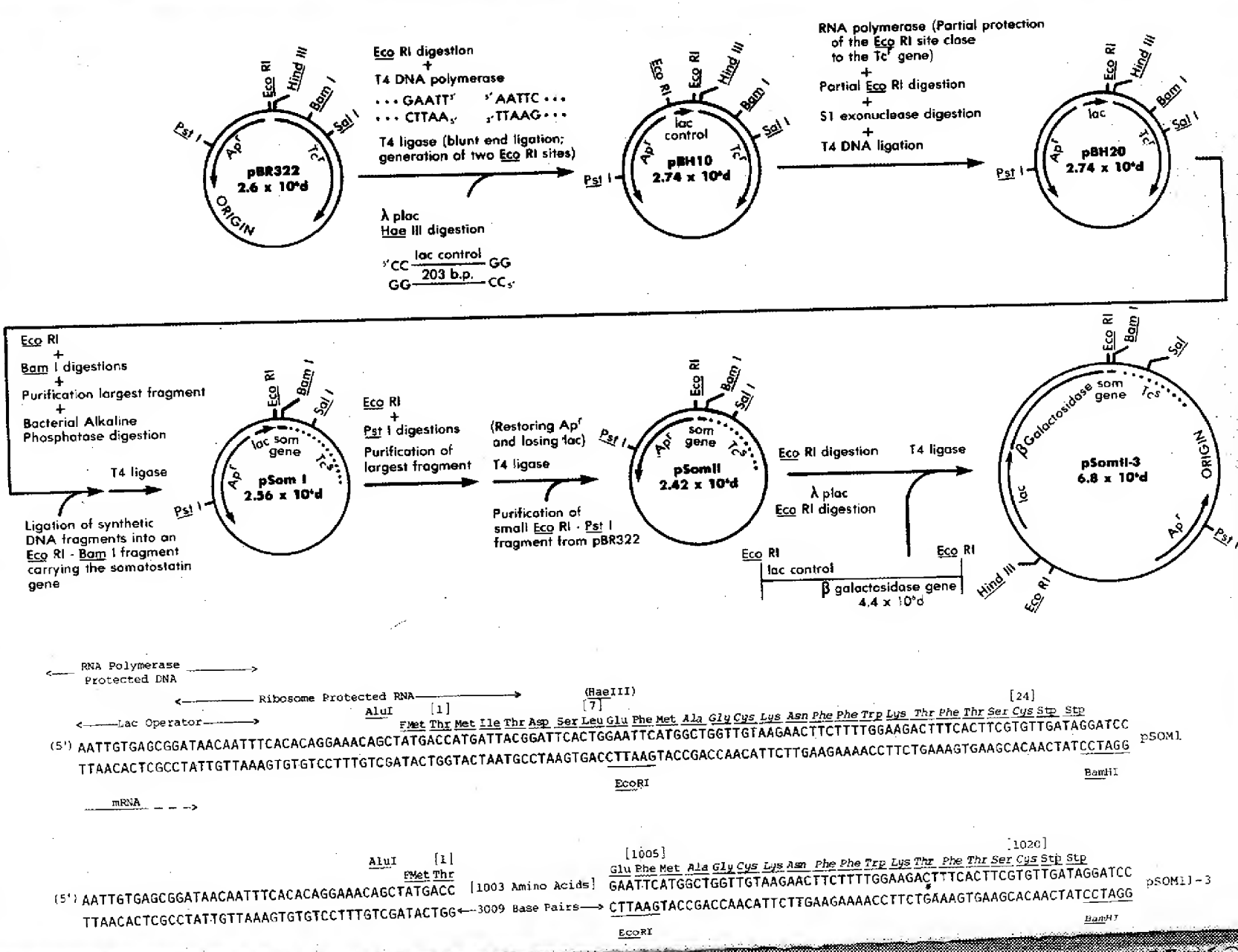


Fig. 2. Chemical synthesis of the somatostatin gene. (a) Eight oligodeoxyribonucleotides, labeled A through H, were synthesized by the modified triester method (*7*, *23*). The codons are indicated, and their corresponding amino acids are given. The eight fragments were designed to have at least five nucleotide complementary overlaps to ensure efficient joining by T4 DNA ligase. (b) Recent improvements in the synthesis of fully protected trimers, which constitute codon blocks and are the basic units for building longer oligodeoxyribonucleotides. With an excess of **1** (2 mmole), the coupling reaction with **2** (1 mmole) went almost to completion in 60 minutes with the aid of a powerful coupling reagent, 2,4,6-triisopropylbenzenesulfonyl tetrazolide (TPSTe, 4 mmole) (*2*). The 5′-protecting group was removed with 2 percent benzene sulfonic acid, and the 5′-hydroxyl dimer **5** could be separated from an excess of 3′-phosphodiester monomer **4** by simple solvent extraction with aqueous $NaHCO_3$ solution in $CHCl_3$. The fully protected trimer block was prepared successively from the 5′-hydroxyl dimer **5**, **1** (2 mmole), and TPSTe (4 mmole) and isolated by chromatography on silica gel (*24*). These improvements simplify the purification step and lead to an increase in the overall yields of trimer blocks and to a decrease in the working time by at least a factor of 2 (*21*). The eight oligodeoxyribonucleotides then were synthesized from the trimers by published procedures (*7*). The final products, after removal of all protecting groups, were purified by high-pressure liquid chromatography on Permaphase AAX (*25*). The purity of each oligomer was checked by homochromatography on thin-layer DEAE-cellulose and also by gel electrophoresis in 20 percent acrylamide (slab) after labeling of the oligomers with [γ-^{32}P]ATP in the presence of polynucleotide kinase. One major labeled product was obtained from each DNA fragment.



lactosidase subunit structure (pSOM11-3).

The plasmid construction scheme (Fig. 3) begins with plasmid pBR322, a well-characterized cloning vehicle (*8*). The lac elements were introduced to this plasmid by insertion of an Hae III restriction endonuclease fragment (203 nucleotides) carrying the lac promoter, catabolite-gene-activator-protein binding site, operator, ribosome binding site, and the first seven amino codons of the β-galactosidase structural gene (*9*) (Figs. 3 and 4). The Hae III fragment was derived from λplac5 DNA. The Eco RI–cleaved pBR322 plasmid, which had its termini repaired with T4 DNA polymerase and deoxyribonucleotide triphosphates, was blunt-end ligated to the Hae III fragment to create Eco RI termini at the insertion points. Joining of these Hae III and repaired Eco RI termini generate the Eco RI restriction site (Figs. 3 and 4) at each terminus. Transformants of *E. coli* RR1 (*8*) with this DNA were selected for resistance to tetracycline (Tc) and ampicillin (Ap) on 5-bromo-4-chloro-indolylgalactoside (X-gal) medium (*10*). On this indicator medium, colonies constitutive for the synthesis of β-galactosidase by virtue of the increased number of lac operators titrating repressor, are identified by their blue color. Two orientations of the Hae III fragment are possible, but these were distinguished by the asymmetric location of an Hha restriction site in the fragment. Plasmid pBH10 was further modified to eliminate the Eco RI endonuclease site distal to the lac operator (pBH20).

The eight chemically synthesized oligodeoxyribonucleotides (Fig. 2) were labeled at the 5′ termini with [γ-^{32}P]ATP (adenosine triphophatase) by T4 polynucleotide kinase and joined with T4 DNA ligase. Through hydrogen bonding between the overlapping fragments, the somatostatin gene self-assembles and eventually polymerizes into larger molecules because of the cohesive restriction site termini. The ligated products were treated with Eco RI and Bam HI restriction endonucleases to generate the somatostatin gene (Fig. 2).

The synthetic somatostatin gene fragment with Eco RI and Bam HI termini was ligated to the pBH20 plasmid, previously treated with the Eco RI and Bam HI restriction endonucleases and alkaline phosphatase. The treatment with alkaline phosphatase provides a molecular selection for plasmids carrying the inserted fragment (*11*). Ampicillin-resistant transformants obtained with this ligated DNA were screened for tetracycline sensitivity, and several were examined for the insertion of an Eco RI–Bam HI fragment of the appropriate size.

Both strands of the Eco RI–Bam HI fragments of plasmids from two clones were analyzed by a nucleotide sequence analysis (*12*) starting from the Bam HI and Eco RI sites. The sequence analysis was extended into the lac-controlling elements; the lac fragment sequence was in-

Fig. 3 (facing page, left). Construction of recombinant plasmids. Plasmid pBR322 was used as the parental plasmid (*8*). Plasmid DNA (5 μg) was digested with the restriction endonuclease Eco RI. The reaction was terminated by extraction with a mixture of phenol and chloroform; the DNA was precipitated with ethanol and resuspended in 50 μl of T4 DNA polymerase buffer (*26*). The reaction was started by the addition of 2 units of T4 DNA polymerase. The reaction (held for 30 minutes at 37°C) was terminated by extraction with phenol and chloroform and precipitation with ethanol. The λplac5 DNA (3 μg) was digested with the endonuclease Hae III (*8*). The digested pBR322 DNA was blunt-end ligated with the Hae III–digested λplac5 DNA in a final volume of 30 μl with T4 DNA ligase (hydroxylapatite fraction) (*27*) in 20 m*M* tris-HCl (*p*H 7.6), 10 m*M* $MgCl_2$, 10 m*M* dithiothreitol, and 0.5 m*M* ATP for 12 hours at 12°C. The ligated DNA mixture was dialyzed against 10 m*M* tris-HCl (*p*H 7.6) and used to transform *E. coli* strain RR1 (*8*). Transformants were selected for tetracycline resistance (Tc^r) and ampicillin resistance (Ap^r) on antibiotic (20 μg/ml) minimal X-gal (40 μg/ml) medium (*10*). Colonies constitutive for the synthesis of β-galactosidase were identified by their blue color. After 45 independently isolated blue colonies were screened, three of them were found to contain plasmids with two Eco RI sites separated by approximately 200 base pairs (*28*). Plasmid pBH10 was shown to carry the fragment in the desired orientation, that is, lac transcription going into the Tc^r gene of the plasmid. Plasmid pBH10 was further modified to eliminate the Eco RI site distal to the lac operator and plasmid pBH20 was obtained (*29*). The nucleotide sequence from the Eco RI site into the lac-control region of pBH20 (data not shown), was confirmed. This plasmid was used for cloning the synthetic somatostatin gene. Plasmid pBH20 (10 μg) was digested with endonucleases Eco RI and Bam HI and treated with bacterial alkaline phosphatase (0.1 unit of BAPF, Worthington), and incubation was continued for 10 minutes at 65°C. The reaction mixtures were extracted with a mixture of phenol and chloroform, and the DNA was precipitated with ethanol (*30*). Somatostatin DNA (50 μl of a solution containing 4 μg/ml) was ligated with the Bam HI–Eco RI, alkaline phosphatase–treated pBH20 DNA in a total volume of 50 μl with the use of 4 units of T4 DNA ligase for 2 hours at 22°C (*31*). In a control experiment, Bam HI–Eco RI alkaline phosphatase–treated pBH20 DNA was ligated in the absence of somatostatin DNA under similar conditions. Both preparations were used to transform *E. coli* RR1. Transformants were selected on minimal X-gal antibiotic plates. Ten Tc^r transformants were isolated. In the control experiment no transformants were obtained. Four out of the ten transformants contained plasmids with both an Eco RI and a Bam HI site. The size of the small Eco RI–Bam HI fragment of these recombinant plasmids was in all four instances similar to the size of the in vitro prepared somatostatin DNA. Base sequence analysis (*12*) revealed that the plasmid pSOM1 had the desired somatostatin DNA fragment inserted (data not shown). Because of the failure to detect somatostatin activity from cultures carrying plasmid pSOM1, a plasmid was constructed in which the somatostatin gene could be located at the COOH-terminus of the β-galactosidase gene, keeping the translation in phase. For the construction of such a plasmid, pSOM1 (50 μg) was digested with restriction enzymes Eco RI and Pst I. A preparative 5 percent polyacrylamide gel was used to separate the large Pst I–Eco RI fragment that carries the somatostatin gene from the small fragment carrying the lac control elements (*12*). In a similar way plasmid pBR322 DNA (50 μg) was digested with Pst I and Eco RI restriction endonucleases, and the two resulting DNA fragments were purified by preparative electrophoresis on a 5 percent polyacrylamide gel. The small Pst I–Eco RI fragment from pBR322 (1 μg) was ligated with the large Pst I–Eco RI DNA fragment (5 μg) from pSOM1. The ligated mixture was used to transform *E. coli* RR1, and transformants were selected for Ap^r on X-gal medium. Almost all the Ap^r transformants (95 percent) gave white colonies (no lac operator) on X-gal indicator plates. The resulting plasmid, pSOM11, was used in the construction of plasmid pSOM11-3. A mixture of 5 μg of pSOM11 DNA and 5 μg of λplac5 DNA was digested with Eco RI. The DNA was extracted with a mixture of phenol and chloroform; the extract was precipitated by ethanol, and the precipitate was resuspended in T4 DNA ligase buffer (50 μl) in the presence of T4 DNA ligase (1 unit). The ligated mixture was used to transform *E. coli* strain RR1. Transformants were selected for Ap^r on X-gal plates containing ampicillin and screened for constitutive β-galactosidase production. Approximately 2 percent of the colonies were blue (such as pSOM11-1 and 11-2). Restriction enzyme analysis of plasmid DNA obtained from these clones revealed that all the plasmids carried a new Eco RI fragment of approximately 4.4 megadaltons, which carries the lac operon control sites and most of the β-galactosidase gene (*13, 14*). Two orientations of the Eco RI fragment are possible, and the asymmetric location of a Hind III restriction in this fragment can indicate which plasmids had transcription proceeding into the somatostatin gene. The clones carrying plasmids pSOM11-3, pSOM11-5, pSOM11-6, and pSOM11-7 contained the Eco RI fragment in this orientation.

Fig. 4 (facing page, right). Nucleotide sequences of the lac-somatostatin plasmids. The nucleotide sequence of the lac control elements, β-galactosidase structural gene, and the synthetically derived somatostatin DNA, are depicted (*9, 14, 27*) along with the restriction endonuclease substrate sites. The nucleotide sequence of pSOM1, as depicted, was confirmed (legends to Figs. 3 and 5). The nucleotide sequence of pSOM11-3 was inferred from published data (*9, 13, 14, 27*). The amino acid sequence of somatostatin is italicized. The amino acid sequence numbers of β-galactosidase are in brackets.

tact, and in one case, pSOM1, the nucleotide sequence of both strands were independently determined, each giving the sequence shown in Fig. 3. In the other case, the sequence was identical except for a base pair deletion (A·T) at a position equivalent to the junction of the B-C oligonucleotides in the original DNA fragment. The basis for the deletion is unclear.

The standard radioimmune assays (RIA) for somatostatin (*2*) were modified by decreasing the assay volume and by using phosphate buffer (Fig. 6). This modification proved suitable for the detection of somatostatin in *E. coli* extracts. Bacterial cell pellets, extracts, or cultures were treated overnight in 70 percent formic acid containing cyanogen bromide (5 mg/ml). Formic acid and cyanogen bromide were removed under vacuum over KOH before the assay. Initial experiments with extracts of *E. coli* strain RR1 (the recipient strain) (*10*) indicated that less than 10 pg of somatostatin could easily be detected in the presence of 16 μg or more of cyanogen bromide–treated bacterial protein. More than 2 μg of protein from formic acid–treated bacterial extracts interfered somewhat by increasing the background, but cyanogen bromide cleavage greatly reduced this interference. Reconstruction experiments showed that somatostatin is stable in cyanogen bromide–treated extracts.

The DNA sequence analysis of pSOM1 indicated that the clone carrying this plasmid should produce a peptide containing somatostatin. However, to date all attempts to detect somatostatin radioimmune activity from extracts of cell pellets or culture supernatants have been unsuccessful. Negative results were also obtained when the growing culture was added directly to 70 percent formic acid and cyanogen bromide. We calculate that *E. coli* RR1 (pSOM1) contains less than six molecules of somatostatin per cell. In a reconstruction experiment we have observed that exogenous somatostatin is degraded very rapidly by *E. coli* RR1 extracts. The failure to find somatostatin activity might be accounted for by intracellular degradation by endogenous proteolytic enzymes.

If the failure to detect somatostatin activity from pSOM1 was due to proteolytic degradation of the small protein (Fig. 4), attachment to a large protein might stabilize it. The β-galactosidase structural gene has an Eco RI site near the COOH-terminus (*13*). The available data on the amino acid sequence of this protein (*13, 14*) suggested that it would be possible to insert the Eco RI–Bam HI somatostatin gene into the site and maintain the proper reading frame for the correct translation of the somatostatin gene (Fig. 4).

The construction of this plasmid is outlined in Fig. 3. The Eco RI–Pst fragment of the pSOM1 plasmid, with the lac-controlling element, was removed and replaced with the Eco RI–Pst fragment of pBR322 to produce the plasmid pSOM11. The Eco RI fragment of λplac5, carrying the lac operon control region and most of the β-galactosidase structural gene, was inserted into the Eco RI site of pSOM11. Two orientations of the Eco RI lac fragment of λplac5 were expected. One of these orientations would maintain the proper reading frame into the somatostatin gene, the other would not.

A number of independently isolated clones (with plasmid designations pSOM11-2 and pSOM11-3) were analyzed for somatostatin activity, as described above. In constrast to the results of experiments with pSOM1, four clones (pSOM11-3, 11-5, 11-6, and 11-7) were

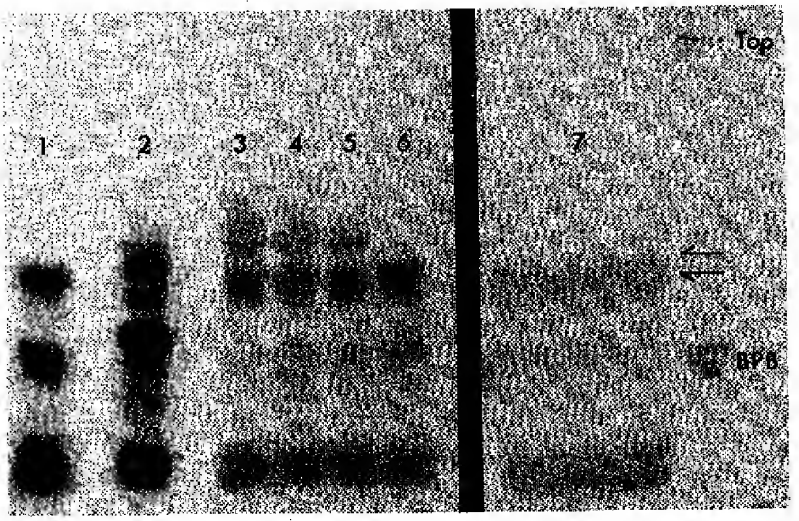


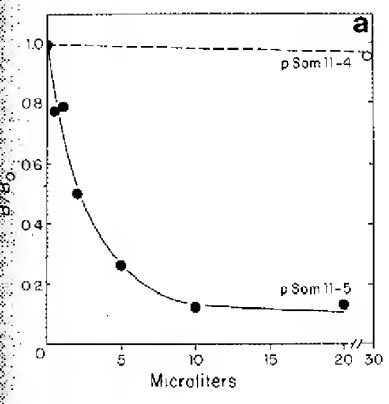
Fig. 5. Ligation and acrylamide gel analysis of somatostatin DNA. The 5'-OH termini of the chemically synthesized fragments A through H (Fig. 2a) were labeled and phosphorylated separately. Just prior to the kinase reaction, 25 μc of [γ-^{32}P]ATP (~ 1500 c/mmole) (*12*) was evaporated to dryness in 0.5-ml Eppendorf tubes. The fragment (5 μg) was incubated with 2 units of T4 DNA kinase (hydroxylapatite fraction, 2500 unit/ml) (*26*), in 70 m*M* tris-HCl, *p*H 7.6, 10 m*M* $MgCl_2$, and 5 m*M* dithiothreitol in a total volume of 150 μl for 20 minutes at 37°C. To ensure maximum phosphorylation of the fragments for ligation purposes, 10 μl of a mixture consisting of 70 m*M* tris-HCl, *p*H 7.6, 10 m*M* $MgCl_2$, 5 m*M* dithiothreitol, 0.5 m*M* ATP, and 2 units of DNA kinase were added, and incubation continued for an additional 20 minutes at 37°C. The fragments (250 ng/μl) were stored at −20°C without further treatment. Kinase-treated fragments A, B, E, and F (1.25 μg each) were ligated in a total volume of 50 μl in 20 m*M* tris-HCl (*p*H 7.6), 10 m*M* $MgCl_2$, 10 m*M* dithiothreitol, 0.5 m*M* ATP, and 2 units of T4 DNA ligase (hydroxylapatite fraction, 400 unit/ml) (*26*), for 16 hours at 4°C. Fragments C, D, G, and H were ligated under similar conditions. Samples (2 μl) were removed for analysis by electrophoresis on a 10 percent polyacrylamide gel and subsequent autoradiography (*16*) (lanes 1 and 2, respectively). The fast migrating material represents unreacted DNA fragments. Material migrating with the bromophenol blue dye (*BPB*) is the monomeric form of the ligated fragments. The slowest migrating material represents dimers, which form by virtue of the cohesive ends, of the ligated fragments A, B, E, and F (lane 1) and C, D, G, and H (lane 2). The dimers can be cleaved by restriction endonuclease Eco RI or Bam HI, respectively (data not shown). The two half molecules (ligated A + B + E + F and ligated C + D + G + H) were joined by an additional ligation step carried out in a final volume of 150 μl at 4°C for 16 hours. A sample (1 μl) was removed for analysis (lane 3). The reaction mixture was heated for 15 minutes at 65°C to inactivate the T4 DNA ligase. The heat treatment does not affect the migration pattern of the DNA mixture (lane 4). Enough restriction endonuclease Bam HI was added to the reaction mixture to cleave the multimeric forms of the somatostatin DNA in 30 minutes at 37°C (lane 5). After the addition of NaCl to a concentration of 100 m*M*, the DNA was digested with Eco RI endonuclease (lane 6). The restriction endonuclease digestions were terminated by phenol-chloroform extraction of the DNA. The somatostatin DNA fragment was purified from unreacted and partially ligated DNA fragments by preparative electrophoresis on a 10 percent polyacrylamide gel. The band indicated with an arrow (lane 7) was excised from the gel, and the DNA was eluted by slicing the gel into small pieces and extracting the DNA with elution buffer (0.5*M* ammonium acetate, 10 m*M* $MgCl_2$, 0.1 m*M* EDTA, and 0.1 percent sodium dodecyl sulfate) overnight at 65°C (*12*). The DNA was precipitated with two volumes of ethanol, centrifuged, redissolved in 200 μl of 10 m*M* tris-HCl (*p*H 7.6), and dialyzed against the same buffer, resulting in a somatostatin DNA concentration of 4 μg/ml.

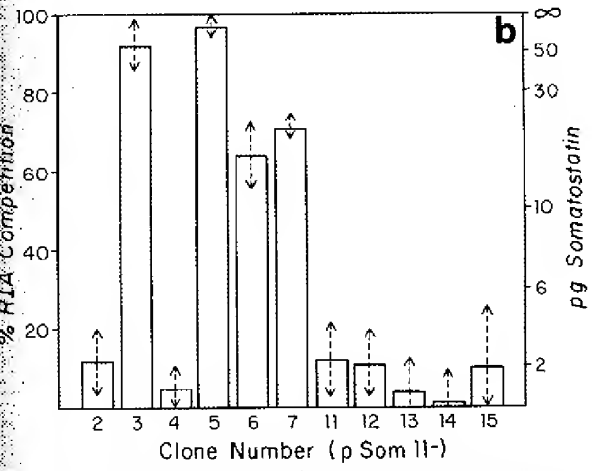
found to have easily detectable somatostatin radioimmune activity (Fig. 6, a and b). Restriction fragment analysis revealed that pSOM11-3, pSOM11-5, pSOM11-6, and pSOM11-7 had the desired orientation of the lac operon, whereas pSOM11-2 and 11-4 had the opposite orientation. Thus, there is a perfect correlation between the correct orientation of the lac operon and the production of somatostatin radioimmune activity.

The design of the somatostatin plasmid predicts that the synthesis of somatostatin would be under the control of the lac operon. The lac repressor gene is not included in the plasmid, and the recipient strain (*E. coli* RR1) contains the wild-type chromosomal lac repressor gene, which produces only 10 to 20 repressor molecules per cell (*15*). The plasmid copy number (and therefore the number of lac operators) is approximately 20 to 30 per cell and complete repression is impossible. The specific activity of somatostatin in *E. coli* RR1 (pSOM11-3) was increased by IPTG, an inducer of the lac operon (Table 1). As expected, the level of induction was low, varying from 2.4- to 7-fold. In experiment 7 (Table 1), the α activity (*14*), a measure of the first 92 amino acids of β-galactosidase, also was induced by a factor of 2.

In several experiments (Table 1 and other experiments not shown), no somatostatin radioimmune activity was detected prior to cyanogen bromide cleavage of the total cellular protein. Since the antiserum used in the radioimmune assay, S39, requires a free NH_2-terminal alanine, no activity was expected prior to cyanogen bromide cleavage. After cleavage by cyanogen bromide, cell extracts were chromatographed on Sephadex G-50 in 50 percent acetic acid (Fig. 6c). In this system, somatostatin is well separated from excluded large peptides and fully included small molecules. Only extracts of clones positive for somatostatin exhibited radioimmune activity in the column fractions, and this activity elutes in the same position as chemically synthesized somatostatin.

The strains carrying the Eco RI lac operon fragment (such as pSOM11-2 and pSOM11-3) segregate with respect to the plasmid phenotype. For example, after





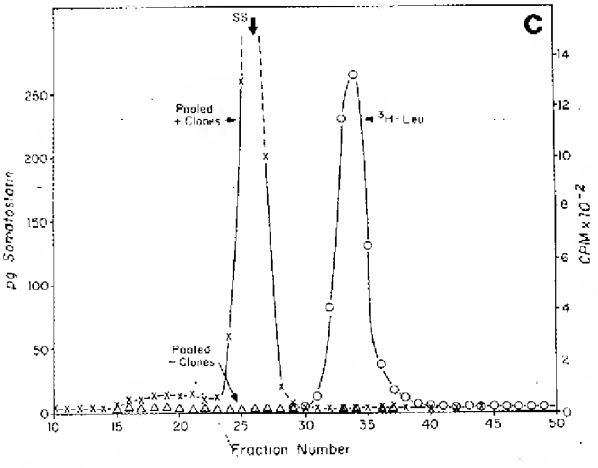


Fig. 6. Radioimmune assay for somatostatin activity. The assay used is a modification of existing methods (*2*). [Tyr[11]]Somatostatin (a gift from W. Vale) was iodinated by a chloramine T procedure (*2*). To assay for somatostatin, the sample, usually in 70 percent formic acid containing cyanogen bromide (5 mg/ml), was dried in a conical polypropylene tube (0.7 ml, Sarstedt) over moist KOH under vacuum. Then 20 μl of PBSA buffer [75 m*M* NaCl; 75 m*M* sodium phosphate (*p*H 7.2; bovine serum albumin (1 μg/ml): and sodium azide (0.2 mg/ml)] was added, followed by 40 μl of a [^{125}I]somatostatin mixture and 20 μl of a 1000-fold dilution in PBSA of rabbit antiserum to somatostatin S39 (*2*) obtained from W. Vale. The [^{125}I]somatostatin mixture contained (per milliliter of PBSA buffer) 250 μg of normal rabbit gamma globulin (Antibodies, Inc.), 1500 units of Trasylol (Calbiochem), and about 100,000 counts of [Tyr[11]-^{125}I]somatostatin. After at least 16 hours at room temperature, 0.333 ml of goat antibody to rabbit gamma globulin (Antibodies, Inc., $P = .03$) in PBSA buffer was added to the sample tubes. The mixture was incubated for 2 hours at 37°C, cooled to 5°C, then centrifuged at 10,000*g* for 5 minutes. The supernatant was removed, and the radioactivity in the pellet was counted in a gamma counter. With the amount of antiserum used, 20 percent of the counts was precipitated with no unlabeled competing somatostatin. The background with infinite somatostatin (200 ng) was usually 3 percent. One-half maximum competition was obtained with 10 pg of somatostatin. (a) Competition by bacterial extracts. Strains *E. coli* RR1 (pSOM11-5) and *E. coli* RR1 (pSOM11-4) were grown at 37°C to 5 × 10^8 cell/ml in L broth. Then IPTG was added to a concentration of 1 m*M* and growth continued for 2 hours. Portions (1 ml) were centrifuged for a few seconds in an Eppendorf centrifuge, and the pellets were suspended in 500 μl of 70 percent formic acid containing cyanogen bromide (5 mg/ml). After approximately 24 hours at room temperature, the samples were diluted tenfold in water, and the indicated volumes were assayed in triplicate for somatostatin. B/B_0 is the ratio of [^{125}I]somatostatin bound in the presence of sample to that bound in the absence of competing somatostatin. Each point is the average of triplicate tubes. The protein content of the undiluted samples were determined to be 2.2 mg/ml for *E. coli* RR1 (pSOM11-5) and 1.5 mg/ml for *E. coli* RR1 (pSOM-4). (b) The initial screening of clones for somatostatin. Cyanogen bromide-treated extracts of 11 clones (such as pSOM11-2 and pSOM11-3) were made as described above for (a). A sample (30 μl) of each extract was taken in triplicate for radioimmune assay. The range of assay points is indicated. The values for picograms of somatostatin were read from a standard curve obtained as part of the same experiment. (c) Gel filtration of cyanogen bromide-treated extracts. Formic acid and cyanogen-treated extracts of the positive clones (11-3, 11-5, 11-6, and 11-7) were pooled (total volume, 250 μl), dried, and resuspended in 0.1 ml of 50 percent acetic acid. [^{3}H]Leucine was added, and the sample was applied to a column (0.7 by 47 cm) of Sephadex G-50 in 50 percent acetic acid. Portions (50 μl) of the column fractions were assayed for somatostatin. Pooled negative clone extracts (11-2, 11-4, and 11-11) were treated identically. On the same column known somatostatin (Beckman Instruments, Inc.) elutes as indicated (*SS*).

Table 1. Somatostatin radioimmune specific activity. Abbreviations: LB, Luria broth, IPTG, isopropylthiogalactoside; CNBr, cyanogen bromide; SS, somatostatin. Protein was measured by the method of Bradford (*32*).

Experiment	Strain	Medium	IPTG 1 m*M*	CNBr 5 mg/ml	SS/protein (pg/mg)
1	11-2	LB	+	+	< 0.1
	11-3	LB	+	+	12
	11-4	LB	+	+	< 0.4
	11-5	LB	+	+	15
2	11-3	LB	+	+	12
	11-3	LB	+	−	< 0.1
3	11-3	LB	+	+	61
	11-3	LB	−	+	8
	11-3	LB	+	−	< 0.1
4	11-3	LB	+	+	71
	11-3	VB + glycerol*	+	+	62
5	11-3	LB + glycerol	+	+	250
6	11-3	LB	+	+	320
	11-2	LB	+	+	< 0.1
7	11-3	LB	+	+	24
	11-3	LB	−	+	10

*Vogel-Bonner minimal medium plus glycerol.

about 15 generations, about one-half of the *E. coli* RR1 (pSOM11-3) culture was constitutive for β-galactosidase, that is, carried the lac operator, and about half of the nonconstitutive colonies were ampicillin-sensitive. Strains positive (pSOM11-3) and negative (pSOM11-2) for somatostatin are unstable, and, therefore, the growth disadvantage presumably comes from the overproduction of the large but incomplete and inactive galactosidase. The yield of somatostatin has varied from 0.001 to 0.03 percent of the total cellular protein (Table 1) probably as the result of the selection for cells in culture having plasmids with a deleted lac region. The highest yields of somatostatin have been from preparations where growth was started from a single Ap-resistant, constitutive colony. Even in these cases, 30 percent of the cells at harvest had deletions of the lac region.

Several moderate scale (up to 10 liters) attempts have been made to purify somatostatin from *E. coli* strain RR1 (pSOM11-3). The initial purification scheme was based on known purification properties of β-galactosidase followed by purification of the cyanogen bromide cleavage products of the chimeric protein. However, essentially all of the somatostatin activity found in the crude extract is insoluble and is found in the pellet from the first low speed centrifugation. The activity can be solubilized in 70 percent formic acid, 6*M* guanidinium hydrochloride, 8*M* urea, or 2 percent sodium dodecyl sulfate. Somatostatin activity has been enriched approximately 100-fold from the cellular debris by cyanogen bromide cleavage, and subsequent alcohol extraction and chromatography on Sephadex G-50 in 50 percent acetic acid.

Recent improvements in the chemical synthesis of DNA provide the opportunity to synthesize quickly DNA with biological interest for genetic manipulation and experimentation. As illustrated earlier (*16, 17*), in vitro recombinant DNA techniques and molecular cloning enhance the experimental value of chemically synthesized DNA. There are two well-established methods for the synthesis of DNA. The phosphodiester method of Khorana and co-workers (*18*) and the more recently developed modified phosphotriester method (*7*). Both methods are capable of producing functional DNA (*16, 17, 19, 20*); however, the triester method is probably faster. Moreover, a method for rapidly synthesizing trimer blocks (codons) as building units for longer oligodeoxyribonucleotides (*21*) (Fig. 2b) has increased the speed of the triester method. From the trimer block library, a hexadecadeoxyribonucleotide now can be obtained in a week. We have established here that the DNA made with this improvement is functional.

The data establishing the synthesis of a polypeptide containing the somatostatin amino acid sequence are summarized as follows. (i) Somatostatin radioimmune activity is present in *E. coli* cells having the plasmid pSOM11-3, which contains a somatostatin gene of proven correct sequence and has the correct orientation of the lac Eco RI DNA fragment. Cells with the related plasmid pSOM11-2, which has the same somatostatin gene but an opposite orientation of the lac Eco RI fragment, produce no detectable somatostatin activity. (ii) As predicted by the design scheme, no detectable somatostatin radioimmune activity is observed until after cyanogen bromide treatment of the cell extract. (iii) The somatostatin activity is under control of the lac operon as evidenced by induction by IPTG, an inducer of the lac operon. (iv) The somatostatin activity cochromatographs with known somatostatin on Sephadex G-50. (v) The DNA sequence of the cloned somatostatin gene is correct. If translation is out of phase, a peptide will be made which is different from somatostatin at every position. Radioimmune activity is detected indicating that a peptide closely related to somatostatin is made, and translation must be in phase. Since translation occurs in phase, the genetic code dictates that a peptide with the exact sequence of somatostatin is made. (vi) Partially purified samples have been independently assayed by W. Vale (Salk Institute). He has confirmed our radioimmune activity with both antiserum S39, which is directed by the NH_2-terminal, and with antiserum S201 which interacts mainly with somatostatin positions 6 through 14. (vii) Finally, the above samples of *E. coli* RR1 (pSOM11-3) extract inhibit the release of growth hormone from rat pituitary cells, whereas samples of *E. coli* RR1 (pSOM11-2) prepared in parallel and with identical protein concentration have no effect on growth hormone release (*22*).

Our results represent the first success in achieving expression (that is, transcription into RNA and translation of that RNA into a protein of a designed amino acid sequence) of a gene of chemically synthesized origin. The large number of plasmid molecules per cell results in a substantial amount (at least 3 percent) of the cellular protein as the β-galactosidase-somatostatin hybrid. This molecule appears to be relatively resistant to endogenous proteolytic activity. There is evidence that abnormally short β-galactosidase peptides are degraded in *E. coli* (*14*) suggesting that the hybrid protein molecule expected from the first somatostatin-lac plasmid (pSOM1) is also rapidly degraded. The synthesis of many gratuitous proteins in *E. coli*, whether large enzymes or smaller polypeptides, may be undetectable for this reason. In cases where the amino acid composition of the protein is appropriate, the precursor technique described here can be employed. This approach could possibly be extended by taking advantage of proteolytic enzymes with amino acid sequence specificity.

The amount of somatostatin synthesized was variable and about a factor of 10 less than the maximum predicted yield. This variability could be interpreted in several ways. Protein degrada-

tion by endogenous proteases, the inability to fully solubilize the chimeric protein, and the selection of altered plasmids could all be contributing factors to the variability in yield. Although recombinant DNA experiments with chemically synthesized DNA are inherently less hazardous than those with DNA from natural sources, consideration should be given to the possible toxicity of the peptide product. A major factor in the choice of somatostatin was its proven low toxicity (*3*). In addition, the experiment was deliberately designed to have the cells produce not free somatostatin but rather a precursor, which would be expected to be relatively inactive. The cloning and growth of cell cultures were performed in a P-3 containment facility.

KEIICHI ITAKURA
TADAAKI HIROSE
ROBERTO CREA
ARTHUR D. RIGGS
Division of Biology,
City of Hope National Medical Center,
Duarte, California 91010
HERBERT L. HEYNEKER*
FRANCISCO BOLIVART
HERBERT W. BOYER
Department of Biochemistry and
Biophysics, University of California,
San Francisco 94143

References and Notes

1. P. Brazeau, W. Vale, R. Burgus, N. Ling, M. Butcher, J. Rivier, R. Guillemin, *Science* **179**, 77 (1973).
2. A. Arimura, H. Sato, D. H. Coy, A. V. Schally, *Proc. Soc. Exp. Biol. Med.* **148**, 784 (1975).
3. W. Vale, N. Ling, J. Rivier, J. Villareal, C. Rivier, C. Douglas, M. Brown, *Metabolism* **25**, 1491 (1976); W. Vale, G. Grant, M. Amoss, R. Blackwell, R. Guillemin, *Endocrinology* **91**, 562 (1972); R. Guillemin and J. E. Gerich, *Annu. Rev. Med.* **27**, 379 (1976); W. Vale, C. Rivier, M. Brown, *Annu. Rev. Physiol.* **39**, 473 (1977).
4. E. Gross and B. Witkop, *J. Am. Chem. Soc.* **83**, 1510 (1961); E. Gross, *Methods Enzymol.* **11**, 238 (1967).
5. W. Fiers *et al.*, *Nature (London)* **260**, 500 (1976).
6. K. Bertrand, L. Korn, F. Lee, T. Platt, C. L. Squires, C. Squires, C. Yanofsky, *Science* **189**, 22 (1975).
7. K. Itakura, N. Katagiri, C. P. Bahl, R. H. Wightman, S. A. Narang, *J. Am. Chem. Soc.* **97**, 7327 (1975); K. Itakura, N. Katagiri, S. A. Narang, C. P. Bahl, R. Wu, *J. Biol. Chem.* **250**, 4592 (1975).
8. F. Bolivar, R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, S. Falkow, *Gene* **2**, 95 (1977).
9. W. Gilbert, J. Gralla, J. Majors, A. Maxam, in *Protein-Ligand Interactions*, H. Sund and G. Blauer, Eds. (De Gruyter, Berlin, 1975), pp. 193–210.
10. J. H. Miller, *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1972).
11. A. Ullrich, J. Shine, J. Chirgwin, R. Pictet, E. Tischer, W. J. Rutter, H. M. Goodman, *Science* **196**, 1313 (1977).
12. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
13. B. Polisky, R. J. Bishop, D. H. Gelfand, *ibid.* **73**, 3900 (1976); P. H. O'Farrell, B. Polisky, D. H. Gelfand, personal communication.
14. A. V. Fowler and I. Zabin, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 1507 (1977); A. I. Bukhari and D. Zipser, *Nature (London) New Biol.* **243**, 238 (1973); K. E. Langley, A. V. Fowler, I. Zabin, *J. Biol. Chem.* **250**, 2587 (1975).
15. W. Gilbert and B. Muller-Hill, *Proc. Natl. Acad. Sci. U.S.A.* **56**, 1891 (1966).
16. H. L. Heyneker *et al.*, *Nature (London)* **263**, 748 (1976).
17. K. Marians, R. Wu, J. Stawinski, T. Hozumi, S. A. Narang, *ibid.*, p. 744.
18. H. G. Khorana, *Pure Appl. Chem.* **17**, 349 (1968).
19. D. V. Goeddel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3292 (1977).
20. H. G. Khorana, personal communication.
21. T. Hirose, R. Crea, K. Itakura, in preparation.
22. W. Vale, personal communication; M. M. Bradford, *Anal. Biochem.* **72**, 248 (1976).
23. Y. Stawinski, T. Hozumi, S. A. Narang, C. P. Bahl, R. Wu, *Nucleic Acids Res.* **4**, 353 (1977).
24. B. T. Hunt and W. Rigby, *Chem. Ind.* (1967), p. 1868.
25. R. A. Henry, J. A. Schmidt, R. C. Williams, *J. Chromatogr. Sci.* **11**, 358 (1973).
26. A. Panet, J. H. van de Sande, P. C. Loewen, H. G. Khorana, A. J. Raae, J. R. Lillehaug, K. Kleppe, *Biochemistry* **12**, 5045 (1973).
27. R. C. Dickson, J. Abelson, W. M. Barnes, W. S. Reznikoff, *Science* **187**, 27 (1975).
28. The position of an asymmetrically located Hha I site in the 203 base pair Hae III lac control fragment (*9*) allows for the determination of the orientation of the Hae III fragment, now an Eco RI fragment, in these plasmids.
29. This was accomplished by preferential Eco RI endonuclease cleavage at the distal site by partial protection with RNA polymerase of the other Eco RI site localized between the Tc' and lac promoters, which are only about 40 base pairs apart. After binding RNA polymerase, the DNA (5 μg) was digested with Eco RI (1 unit) in a final volume of 10 μl for 10 minutes at 37°C. The reaction was stopped by heating at 65°C for 10 minutes. The Eco RI cohesive termini were digested with S1 nuclease in a solution of 25 m*M* sodium acetate (p*H* 4.5), 300 m*M* NaCl, and 1 m*M* $ZnCl_2$ at 25°C for 5 minutes. The reaction mixture was stopped by the addition of EDTA (10 m*M*, final) and tris-HCl (p*H* 8) (50 m*M* final). The DNA was extracted with phenol-chloroform, precipitated with ethanol, and resuspended in 100 μl of T4 DNA ligation buffer. The T4 DNA ligase (1 μl) was added and the mixture was incubated at 12°C for 12 hours. The ligated DNA was transformed in *E. coli* strain RR1, and Ap^r Tc^r transformants were selected on X-gal–antibiotic medium. Restriction enzyme analysis of DNA screened from ten isolated blue colonies revealed that these clones carried plasmid DNA with one Eco RI site. Seven of these colonies had retained the Eco RI site located between the lac and Tc^r promoters.
30. The alkaline phosphatase treatment effectively prevents self-ligation of the Eco RI–Bam HI treated pBH20 DNA, but circular recombinant plasmids containing somatostatin DNA can still be formed upon ligation. Since *E. coli* RR1 is transformed with very low efficiency by linear plasmid DNA, the majority of the transformants will contain recombinant plasmids (*1*).
31. After 10, 20, and 30 minutes, additional somatostatin DNA (40 ng) was added to the reaction mixture (the gradual addition of somatostatin DNA may favor ligation to the plasmid over self-ligation). Ligation was continued for 1 hour and then the mixture was dialyzed against 10 m*M* tris-HCl (p*H* 7.6).
32. M. M. Bradford, *Anal. Biochem.* **72**, 248 (1976).
33. Supported by contracts from Genentech, Inc. to the City of Hope National Medical Center and the University of California. H.W.B. is an investigator of the Howard Hughes Medical Research Institute. We thank Li Shively, Y. Lu, L. Shih, and L. Directo for their assistance in various aspects of the project, and R. A. Swanson for his assistance and encouragement throughout the design and execution of the project. We also thank D. Gelfand and P. O'Farrell for discussing their unpublished data with us.

* Present address: Department of Molecular Genetics, University of Leiden, Wassenaarseweg 64, Leiden, Netherlands.
† Present address: Departamento de Biologia Molecular, Instituto de Investigaciones Biomedicas, Universidad Nacional Autonoma de Mexico, Mexico 20 D.F. Apdo Postal 70228.

2 November 1977

Cytidine 3',5'-Monophosphate (Cyclic CMP) Formation in Mammalian Tissues

Abstract. *Mammalian tissues possess the capacity to synthesize cytidine 3',5'-monophosphate (cyclic CMP) via the enzymatic conversion of cytidine 5'-triphosphate to cyclic CMP by cytidylate cyclase. Cyclic CMP formation occurs best in the presence of manganese or iron, at neutral* p*H, at 37°C, in the absence of detergents, and with whole tissue homogenate fractions. Thus, mammalian tissues are capable of synthesizing not only cyclic AMP and cyclic GMP, but also cyclic CMP.*

Adenosine 3',5'-monophosphate (cyclic AMP) and guanosine 3',5'-monophosphate (cyclic GMP) are purine cyclic nucleotides that are generally thought to influence or regulate numerous cell functions and biological events. In many instances, however, alterations in cell function cannot be accounted for by corresponding or concomitant alterations in the tissue concentrations of either of the two purine cyclic nucleotides. Therefore, the existence of other endogenous regulatory molecules is constantly being sought. Cytidine 3',5'-monophosphate (cyclic CMP) was first identified in cells (leukemia L-1210) by Bloch, who demonstrated also that the addition of exogenous cyclic CMP to L-1210 cells in culture abolishes the characteristic temperature-dependent lag phase and stimulates the resumption of growth or proliferation of these leukemic cells (*1*, *2*). These experimental findings suggest that cyclic CMP, a pyrimidine cyclic nucleotide, may play a biologic role in the control of proliferation of leukemic cells.

Shortly after the discovery of the natural occurrence of cyclic CMP in certain leukemic cells, an enzyme system capable of forming cyclic CMP from its naturally occurring substrate was found in murine myeloid leukemic tumors and in normal mouse liver and spleen (*3*). Thus, our experimental findings on the capacity of mammalian tissues to synthesize cyclic CMP support those of Bloch on the identification of cyclic CMP in malignant cells.

The properties and biologic importance of cytidylate cyclase in normal and malignant mammalian tissues were recently reported briefly (*4*). At the same time Cailla and Delagge reported on the

TAB K

Proc. Natl. Acad. Sci. USA
Vol. 75, No. 8, pp. 3727–3731, August 1978
Biochemistry

A bacterial clone synthesizing proinsulin

(rat preproinsulin/cDNA cloning/solid-phase radioimmunoassay/DNA sequence/fused proteins)

LYDIA VILLA-KOMAROFF*, ARGIRIS EFSTRATIADIS*, STEPHANIE BROOME*, PETER LOMEDICO*, RICHARD TIZARD*, STEPHEN P. NABER†, WILLIAM L. CHICK†, AND WALTER GILBERT*

* Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138; and † Elliot P. Joslin Research Laboratory, Harvard Medical School, and the Peter Bent Brigham Hospital, Boston, Massachusetts 02215

Contributed by Walter Gilbert, June 9, 1978

ABSTRACT **We have cloned double-stranded cDNA copies of a rat preproinsulin messenger RNA in *Escherichia coli* χ1776, using the unique *Pst* endonuclease site of plasmid pBR322 that lies in the region encoding amino acids 181–182 of penicillinase. This site was reconstructed by inserting the cDNA with an oligo(dG)-oligo(dC) joining procedure. One of the clones expresses a fused protein bearing both insulin and penicillinase antigenic determinants. The DNA sequence of this plasmid shows that the insulin region is read in phase; a stretch of six glycine residues connects the alanine at position 182 of penicillinase to the fourth amino acid, glutamine, of rat proinsulin.**

Can the structural information for the production of a higher cell protein be inserted into a plasmid in such a way as to be expressed in a transformed bacterium? To attack this problem, we used as a model rat insulin, an interesting protein that can be identified by immunological and biological means.

Although mature insulin contains two chains, A and B, it is the product of a single longer polypeptide chain. The hormone is initially synthesized as a preproinsulin structure (1, 2). A hydrophobic leader sequence of 23 amino acids at the amino terminus of the nascent chain is cleaved off, presumably as the polypeptide chain moves through the endoplasmic reticulum (2–4), producing a proinsulin molecule. The proinsulin chain folds up and then the C peptide is cleaved from its middle (5). Thus each of the two (nonallelic) insulin genes in the rat (6–8) encodes a polypeptide 109 amino acids long, whose initial structure is NH_2—leader sequence—B chain—C peptide—A chain.

Ullrich *et al.* (9) have cloned double-stranded cDNA copies of rat preproinsulin mRNA isolated from pancreatic islets and determined sequences covering much of those two genes. We have made double-stranded cDNA copies of mRNA from a rat insulinoma (10) and cloned these in the *Pst* (*Providencia stuartii* endonuclease) site of pBR322 (11), which lies within the penicillinase gene.

The *Escherichia coli* penicillinase is a periplasmic protein, the gene for which was recently sequenced (12). Penicillinase is synthesized as a preprotein with a 23 amino acid leader sequence (12, 13), which presumably serves as a signal to direct the secretion of the protein to the periplasmic space, and is removed as the protein traverses the membrane. Insertion of the structural information for insulin into the penicillinase gene should cause expression of the insulin sequence as a fusion product transported outside the cell.

MATERIALS AND METHODS

Bacterial Strains. *E. coli* K-12, strain HB101 [*hsm*$^-$, *hrs*$^-$, *recA*$^-$, *gal*$^-$, *pro*$^-$, *str*^{r} (14)] was initially obtained from H. Boyer. *E. coli* K-12 strain χ1776 (15) (F$^-$, *tonA53, dapD8, minA1, supE42,* Δ40[*gal-uvrB*], λ$^-$, *minB2, rfb-2, nalA25, oms-2, thyA57, metC65, oms-1,* Δ29[*bioH-asd*], *cycB2, cycA1, hsdR2*) was provided by R. Curtiss.

DNA and Enzymes. pBR322 DNA, a gift from A. Poteete, was used to transform *E. coli* HB101. Plasmid DNA was purified according to the procedure of Clewell (16). Avian myeloblastosis virus reverse transcriptase (RNA-dependent DNA polymerase), *E. coli* DNA polymerase I, and terminal transferase were gifts from T. Papas, M. Goldberg, and J. Wilson, respectively. Restriction enzymes were purchased from Bethesda Research Labs and New England BioLabs.

RNA Purification. An x-ray-induced, transplantable rat beta cell tumor (10) was used as source of preproinsulin mRNA. Tumor slices (20 g per preparation) were homogenized, and a cytoplasmic RNA (about 2 mg/g of tissue) was purified from a postnuclear supernatant by Mg^{2+} precipitation (17), followed by extraction with phenol and chloroform, and enriched for poly(A)-containing RNA by oligo(dT)-cellulose chromatography (18). About 4% of the material binds to the column (data from eight preparations). Further purification of the oligo(dT)-cellulose-bound material by sucrose gradient centrifugation and/or polyacrylamide gel electrophoresis showed that the preproinsulin mRNA was a minor component of the preparation.

Double-Stranded cDNA Synthesis. Oligo(dT)-cellulose-bound RNA was used directly as template for double-stranded cDNA synthesis (19), except that a specific p(dT)$_8$dG-dC primer (Collaborative Research) was utilized for reverse transcription. The concentrations of RNA and primer were 7 mg/ml and 1 mg/ml, respectively. All four [α-^{32}P]dNTPs were at 1.25 mM (final specific activity 0.85 Ci/mmol). The reverse transcript was 2% of the input RNA, and 25% of it was finally recovered in the double-stranded DNA product.

Construction of Hybrid DNA Molecules. pBR322 DNA (5.0 μg) was linearized with *Pst*, and approximately 15 dG residues were added per 3′ end by terminal transferase at 15° in the presence of 1 mM Co^{2+} (20) and autoclaved gelatin at 100 μg/ml. Similarly, dC residues were added to 2.0 μg of double-stranded cDNA, which was then electrophoresed in a 6% polyacrylamide gel. Following autoradiography, molecules in the size range of 300 to 600 base pairs (0.5 μg) were eluted from the gel (21). Size selection was done after tailing rather than before because previous experience had indicated that occasionally impurities contaminating DNA extracted from gels inhibits terminal transferase. The eluted double-stranded cDNA was concentrated by ethanol precipitation, redissolved in 10 mM Tris·HCl at pH 8, mixed with 4 μg of dG-tailed pBR322, and dialyzed versus 0.1 M NaCl/10 mM EDTA/10 mM Tris, pH 8. The mixture (4 ml) was then heated at 56° for 2 min, and annealing was performed at 42° for 2 hr. The hybrid DNA was used to transform *E. coli* χ1776.

Transformation and Identification of Clones. Transformation of *E. coli* χ1776 (an EK2 host) with pBR322 (an EK2

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

vector) was performed in a biological safety cabinet in a P3 physical containment facility in compliance with NIH guidelines for recombinant DNA research published in the *Federal Register* [(1976) **41**, 27902–27943].

χ1776 was transformed by a transfection procedure (22) adapted to χ1776 by A. Bothwell (personal communication) and slightly modified as follows: χ1776 was grown in L broth (23) supplemented with diaminopimelic acid at 10 μg/ml and thymidine (Sigma) at 40 μg/ml to OD_{590} of 0.5. Cells (200 ml) were sedimented at 500 × *g* and resuspended by swirling in 1/10th vol of cold buffer containing 70 mM $MnCl_2$, 40 mM Na acetate at pH 5.6, 30 mM $CaCl_2$, and kept on ice for 20 min. The cells were repelleted and resuspended in 1/30th of the original volume in the same buffer. Two milliliters of the annealed DNA preparation was added to the cells. Aliquots of this mixture (0.3 ml) were placed in sterile tubes and incubated on ice for 60 min. The cells were then placed at 37° for 2 min. Broth was added to each tube (0.7 ml) and the tubes were incubated at 37° for 15 min; 200 μl of the cells was spread on sterile nitrocellulose filters (Millipore) overlaying agar plates containing tetracycline at 15 μg/ml. (The filters were boiled to remove detergents before use.) The plates were incubated at 37° for 48 hr. Replicas of the filters were made by a procedure developed by D. Hanahan (personal communication): The nitrocellulose filters containing the transformants were removed from the agar and placed on a layer of sterile Whatman filter paper. A new sterile filter was placed on top of the filter containing the colonies and pressure was applied with a sterile velvet cloth and a replica block. A sterile needle was used to key the filters. The second filter was placed on a new agar plate and incubated at 37° for 48 hr. The colonies on the first filter were screened by the Grunstein–Hogness technique (24), using as probe an 80-nucleotide-long fragment produced by *Hae* III digestion of high specific activity cDNA (9). Positive colonies were rescreened by hybrid-arrested translation (25) as described in the legend of Table 1.

Radioimmunoassays. Two-site solid-phase radioimmunoassays were performed (28). Cells from colonies to be tested were transferred with an applicator stick onto 1.5% agarose containing 30 mM Tris·HCl, pH 8, lysozyme at 0.5 mg/ml, and 10 mM EDTA; released antigen was adsorbed to an IgG-coated polyvinyl disk during a 1-hr incubation at 4°. The wash buffer contained streptomycin sulfate at 300 μg/ml and normal guinea pig serum (Grand Island Biological Co.) instead of normal rabbit serum. Guinea pig antiserum to bovine insulin was purchased from Miles Laboratories.

Standard (liquid) radioimmunoassays were performed using the back titration procedure employing alcohol precipitation of insulin–antibody complexes (29).

DNA Sequencing. DNA sequencing was performed as described by Maxam and Gilbert (30).

RESULTS

Construction and Identification of cDNA Clones. We isolated poly(A)-containing RNA from a transplantable rat insulinoma. This preparation contained preproinsulin mRNA, because it directed the synthesis in a cell-free system of a product precipitable with anti-insulin antibody (data not shown). However, the mRNA yield after further purification was not sufficient for cloning, and therefore we decided to clone cDNA synthesized from the total preparation. In an attempt to enrich the reverse transcript for insulin sequences, we utilized the DNA sequence reported by Ullrich *et al.* (9) to choose a specific primer, $(dT)_8dG$-dC. The product of double-stranded cDNA synthesis (19) was extended by a short oligo(dC) tail about 15 nucleotides in length, and sized on a polyacrylamide gel. A broad size cut averaging 500 base pairs was selected in order to enrich for full-length sequences. We inserted these molecules into the *Pst* site of pBR322 after elongating the 3′-terminal extension of the cleavage site with oligo(dG). We used this oligo(dG)-oligo(dC) joining procedure in order to reconstruct the *Pst* recognition sequence (ref. 31; W. Rowenkamp and R. Firtel, personal communication); approximately 40% of the inserts were excisable with *Pst* after cloning. From about 0.25 μg of tailed cDNA we obtained 2355 transformants in *E. coli* strain χ1776. To identify clones containing insulin sequences, we first screened one-third of the transformants, using as a probe an 80-nucleotide-long *Hae* III fragment of cDNA synthesized from oligo(dT)-bound RNA because the results of Ullrich *et al.* (9) suggested that such a fragment should be insulin specific. About 20% of the clones were positive, but restriction analysis of plasmid DNA from a few candidates showed that the inserts were not insulin sequences. We concluded that our probe was not pure and rescreened some of the positive clones, using hybrid-arrested translation (25). This method is based on the principle that mRNA in the form of an RNA·DNA hybrid does not direct cell-free protein synthesis. We incubated aliquots of oligo(dT)-bound RNA with linearized

Table 1. Hybrid-arrested translation and immunoprecipitation of the cell-free products

Source of arresting DNA	Radioactivity, cpm/20 μl: Acid insoluble	Immunoprecipitable: − Insulin	Immunoprecipitable: + Insulin	% Immunoprecipitable*
Control I (−DNA, −RNA)†	2,570			
Control II (−DNA, +RNA)‡	35,700	12,300	310	36.2
pBR322	28,800	7,850	245	29.0
Clone 3	15,100	3,630	264	26.9
Clone 13	19,600	5,190	350	28.4
Clone 15	18,600	4,850	252	28.7
Clone 16	29,200	8,830	247	32.2
Clone 17	24,000	6,700	316	30.0
Clone 18	15,900	3,690	251	25.8
Clone 19	8,650	587	277	5.0
Clone 20	15,100	4,070	231	30.6
Clone 21	21,100	5,170	223	26.7

Plasmid DNA (about 3 μg) was digested with *Pst*, precipitated with ethanol, and dissolved directly in 20 μl of deionized formamide. After heating for one minute at 95° each sample was placed on ice. Following the addition of 1.5 μg of oligo(dT)-cellulose-bound RNA, piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (Pipes) at pH 6.4 to 10 mM, and NaCl to 0.4 M, the mixtures were incubated for 2 hr at 50°. They were then diluted by the addition of 75 μl of H_2O and ethanol precipitated in the presence of 10 μg of wheat germ tRNA, washed with 70% (vol/vol) ethanol, dissolved in H_2O, and added to a wheat germ cell-free translation mixture (26) containing 10 μCi of [^{3}H]leucine (60 Ci/mmol). Fifty-microliter reaction mixtures were incubated at 23° for 3 hr and then duplicate 2-μl aliquots were removed for trichloroacetic acid precipitation. From the remainder two 20-μl aliquots were treated with ribonuclease, diluted with immunoassay buffer, and analyzed for the synthesis of immunoreactive preproinsulin by means of a double antibody immunoprecipitation (27) in the absence or presence of 10 μg of bovine insulin. The washed immunoprecipitates were dissolved in 1 ml of NCS (Amersham) and assayed in 10 μl of Omnifluor (New England Nuclear) by liquid scintillation counting.

* Calculated using the formula [(immunoprecipitable radioactivity in the absence of insulin) − (immunoprecipitable radioactivity in the presence of insulin)]/[(acid-insoluble radioactivity) − (acid-insoluble radioactivity of control I)].

† Reaction mixture incubated in the absence of added RNA.

‡ Cell-free translation by the direct addition of oligo(dT)-cellulose-bound RNA into the reaction mixture.

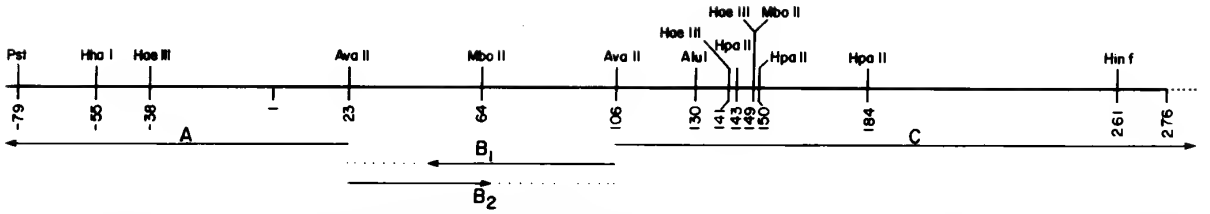


FIG. 1. Restriction map of the insertion in clone pI19. Each restriction site is identified by a number indicating the 5′-terminal nucleotide generated by cleavage at the message strand. Nucleotides are numbered beginning with the first base of the sequence encoding proinsulin. Nucleotides in the 5′ direction from position 1 in the message strand are identified by negative numbers, beginning with −1. Arrows indicate the sequenced fragments; those pointing to the left indicate sequences derived from the antimessage strand, and those pointing to the right indicate sequences derived from the message strand. The uniquely labeled restriction fragments were generated as follows: Following excision with *Pst*, DNA of the insertion was digested with *Ava* II and end labeled. Fragments A and C purified from a polyacrylamide gel were sequenced directly because the *Pst* ends do not label significantly. Fragment B was strand separated on a polyacrylamide gel and sequenced in both directions. The exact number of C·G pairs in the right-hand tail before the *Pst* site could not be counted.

DNA from nine clones under conditions favoring DNA·RNA hybridization (32), added them to cell-free translation systems, and assayed for a specific inhibition of insulin synthesis. Table 1 shows that one of the plasmids, pI19, inhibited the synthesis of immunoprecipitable material. Restriction endonuclease digestions of the *Pst*-excised insert of pI19 with several enzymes generated fragments whose sizes were consistent with the sequence of Ullrich *et al.* (9). We confirmed the presence of insulin DNA in pI19 by direct DNA sequence analysis and screened the rest of the clones with purified pI19 insert labeled by nick translation. About 2.5% (48/1745) of the clones hybridized strongly to this probe. There must have been enrichment for insulin sequence at some step of our procedure, because hybridization analysis using cloned insulin DNA as probe showed the presence of only 0.3% insulin mRNA in the original oligo(dT)-bound RNA.

Sequence Information. Fig. 1 shows the restriction map of the insertion in clone pI19 and Fig. 2 shows the sequence of the insert. It corresponds to rat insulin I (5, 33) and encodes the entire preproinsulin chain with the exception of the first two amino acid residues of the reported preregion (1). It therefore extends the sequence determined by Ullrich *et al.* (9) by twenty-five 5′-terminal nucleotides. It also verifies the reported amino acid residues for positions −14, −17, −18, and −20; it identifies the previously uncertain residue −15; and it identifies the unknown residue −19. However, the residues at positions −16 and −21 differ from those reported (1).

The sequence deviates from that determined by Ullrich *et al.* (9) at the region immediately after the UGA terminator, where a GAGTC sequence occurs, predicting a *Hin*f cleavage site that we have experimentally verified. Furthermore, only moderate agreement exists between the two sequences for the next 15 nucleotides of the 3′ untranslated region.

Expression. Almost two-thirds of the clones carrying inserts were ampicillin resistant; thus the active site of penicillinase must lie between amino acid residues 23 and 182 (12). The degree of resistance was variable, suggesting the expression of different sequences from the inserts in the form of fused translation products, probably differing in length and stability.

We therefore screened colonies of the 48 clones containing insulin sequence for the presence of insulin antigenic determinants, using a solid-phase radioimmunoassay (28). Polyvinyl sheets coated with antibody molecules will bind specific antigens released from bacteria. The immobilized antigen can then be detected by autoradiography following exposure of the sheets to ^{125}I-labeled antibody. This method permits detection of as little as 10 pg of insulin in a colony. We coated plastic disks with anti-insulin antibody and used ^{125}I-labeled anti-insulin to detect solely insulin antigenic determinants. Disks coated with anti-penicillinase antibody and exposed to ^{125}I-anti-insulin detect the presence of a fused protein, as do disks coated with anti-insulin and exposed to radioiodinated anti-penicillinase.

One clone, pI47, gave positive responses with all of the combinations described above; this indicates the presence of a penicillinase–insulin hybrid polypeptide. Fig. 3 shows some of the results. To determine whether this fused protein is secreted, we grew clone pI47 in liquid culture and extracted the proteins in the periplasmic space by osmotic shock, a method that does not lyse bacteria (34). Fig. 4 shows that the insulin

FIG. 2. DNA sequence of the insertion in clone pI19. Nucleotides are numbered using the convention described in Fig. 1. Accordingly, amino acids are numbered beginning with the first amino acid of proinsulin, while the last amino acid of the leader sequence (pre region) is numbered as −1. Restriction endonuclease cleavage sites experimentally verified are underlined and identified. The arrows indicate, in order, the ends of the leader sequence and the peptides B, C, and A. Two nucleotides indicated by double underlining are uncertain.

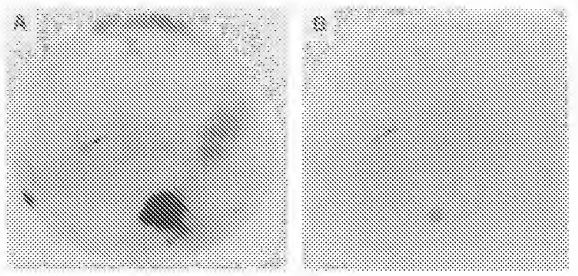


FIG. 3. Initial detection of penicillinase-insulin hybrid polypeptides in an insulin cDNA clone. Cells from colonies of the 48 insulin cDNA clones and from control colonies, χ1776 and χ1776-pBR322, were applied to an agarose/lysozyme/EDTA plate. Positive controls, 5 ng of insulin and 5 ng of penicillinase, each in 1 μl of wash buffer, also were spotted on plate. Antigen was adsorbed to an IgG-coated polyvinyl disk during a 1-hr incubation at 4°. Immobilized antigen was labeled by setting the plastic disk on a solution containing radioiodinated anti-insulin IgG. The autoradiographs are of disks precoated with anti-insulin IgG (*A*) or anti-pencillinase IgG (*B*), exposed on Kodak X-Omat R film using a Du Pont Cronex Lightning Plus intensifying screen for 12 hr at −70°. The arrows indicate the signal generated by clone pI47. The large exposed area in the lower right of (*A*) is the positive control for insulin detection.

antigen was recovered in the distilled water wash of the shock procedure. Table 2 shows that the insulin antigen in the wash is also detectable and quantifiable by a standard radioimmunoassay. The yield of antigen depended on the growth medium; antigen was released by cells grown in M9/glucose/amino acids medium but not by cells grown in brain/heart infusion. We estimate a recovery of about 100 molecules per cell.

Structure of the Fused Protein. We sequenced pI47 to determine the sequence around the junctions. Fig. 5 shows that a proinsulin I cDNA lies in the *Pst* site in the correct orientation and in phase, so that a fused protein can be synthesized. In pI19, the insert is in the correct orientation, but not in phase. In pI47 the oligo(dG)·oligo(dC) region encodes six glycines that connect the penicillinase sequence, ending at amino acid 182 (alanine), to the fourth amino acid (glutamine), of the proinsulin sequence. The cDNA sequence in pI47 extends 26 base pairs past the UGA terminator. Thus, we infer the structure of the fused protein to be penicillinase(24–182)-$(Gly)_6$-proinsulin(4–86).

DISCUSSION

The coding regions of eukaryotic structural genes are often interrupted by introns (35–38), whose transcripts are spliced out of the mature mRNA. Because prokaryotes do not appear to process their messengers, double-stranded cDNA made from a mature messenger is the material of choice to carry eukaryotic structural information into bacteria.

By using cDNA cloning technology and an extremely sensitive method to assay expression, we were able to construct a derivative of *E. coli* strain χ1776 carrying an insulin gene sequence and to detect the synthesis and secretion into the periplasmic space of a fused protein carrying antigenic determinants of both insulin and penicillinase. This was accomplished simply by inserting double-stranded cDNA carrying the structural information for insulin into a restriction site within the structural gene for penicillinase. Not only is the fused DNA sequence expressed as a chain of amino acids, but also the polypeptide folds so as to reveal insulin antigenic shapes. Thus we expect soon to be able to demonstrate biological function for this, or for a similar, fused protein.

We anticipate that the joining of cDNA sequences to nucleotides that lie ahead of the *Pst* site in the penicillinase gene will also produce fused and secreted molecules. Moreover, if the fusion replaces the preproinsulin leader with that of penicillinase it is likely that the new protein will also be secreted by the *E. coli* cell and may even be correctly matured by cleavage of the leader sequence.

Table 2. Immunoreactive insulin concentration in distilled water wash of osmotic shock procedure

Exp.	Insulin, μunits/ml	Cells/ml
1	318	1.5×10^{10}
2	166	6.0×10^{9}
3	386	4.2×10^{10}

Duplicate 0.1-ml aliquots of each sample prepared as described in the legend to Fig. 4 were assayed (29) in a final volume of 0.4 ml using rat insulin standard, a gift from J. Schlichtkrull. One unit = 48 μg. The NaCl/Tris wash, the 20% sucrose wash, and the media of χ1776-pI47 as well as the water wash from osmotic shock of χ1776-pBR322 gave values below the sensitivity of the assay (25 μunits/ml).

Clearly, we have exploited a general method that should lead to the expression and secretion of any eukaryotic protein provided another protein, such as penicillinase, will serve as a carrier, by virtue of its leader sequence. Moreover, the secretion of the eukaryotic protein sequence to the periplasm or extracellular space will both permit its harvest in a purified form and probably eliminate intracellular sources of instability.

Often just an expression of antigens is the goal. In a "shotgun" screening, the existence of a fused protein antigen could be used to identify transformants carrying desired eukaryotic gene fragments. On the other hand, the insertion of a DNA fragment coding for surface antigenic determinants of a virus into a carrier protein should lead to the secretion of a fused protein that could serve as a vaccine, even though no entirely correct virus product is ever produced.

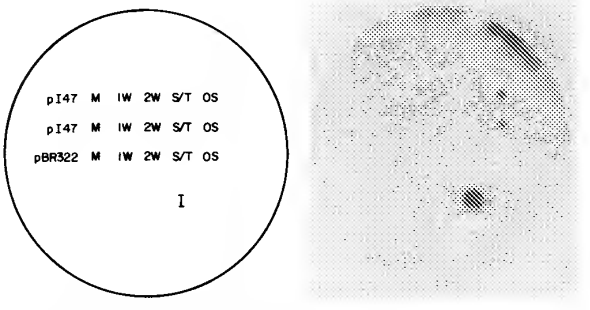


FIG. 4. Release of insulin antigen from χ1776-pI47 cells by osmotic shock. One liter of χ1776-pI47 cells growing at 37° in M9 medium supplemented with 1 g of tryptone, 0.5 g of yeast extract, and 0.5% glucose was harvested at a density of 5×10^7 cells per ml and washed two times in 10 ml of cold 10 mM Tris·HCl, pH 8/30 mM NaCl. The cells were then osmotically shocked (34) in the following manner: The final wash pellet was resuspended in 10 ml of 20% sucrose per 30 mM Tris-HCl, pH 8, at room temperature, made 1 mM in EDTA, shaken at room temperature for 10 min, centrifuged out, resuspended in 10 ml of cold distilled water, shaken in an ice bath for 10 min, and again pelleted. The resulting supernatant was termed the "water wash." As a control, 1 liter of χ1776-pBR322 was grown and treated in a similar manner. Aliquots (1 μl) of each fraction to be assayed for the presence of insulin antigen were applied to the surface of a 1.5% agar plate. (*A*) Positions of each fraction on the plate. M, medium; 1W, first wash supernatant; 2W, second wash supernatant; S/T, sucrose/Tris supernatant; OS, distilled water wash; I, insulin. (*B*) Autoradiograph showing results of a two-site radioimmunoassay of these fractions. Antigen was adsorbed to a polyvinyl disk and labeled by using anti-insulin IgG. The labeled areas correspond to the water washes and the positive control (1 ng insulin). A spectrophotometric assay for β-galactosidase (23) indicated that no more than 4% of cells lyse during this procedure.

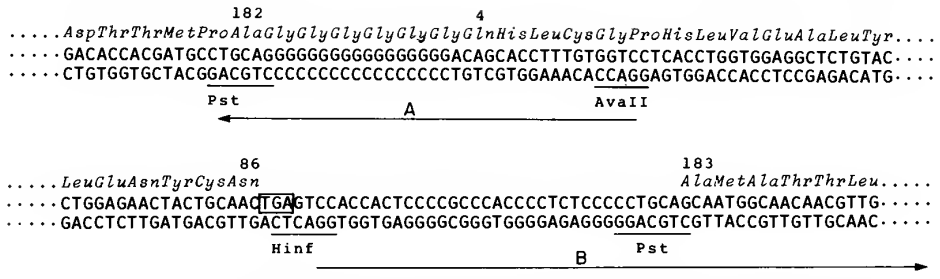


FIG. 5. Partial DNA sequence of the insertion in clone pI47. Clone pI47 DNA was digested with *Hinf* and two fragments, H1 and H2 (ca. 1700 and 280 base pairs long, respectively) were isolated. H1 contains the amino-terminal portion of the penicillinase gene and the bulk of the cDNA insert. H1 was digested with *Ava* II, end labeled, and digested again with *Pst*. A fragment 39 nucleotides long (fragment A, arrow) was isolated and sequenced. Fragment H2 was end labeled and digested with *Alu* I (which cuts at the region corresponding to amino acid 200 of penicillinase). A fragment 88 base pairs long (fragment B, arrow) was isolated and sequenced. The termination sequence TGA is boxed.

We thank David Baltimore, Philip Sharp, and Salvador Luria for the use of the Massachusetts Institute of Technology P3 laboratory. We thank Macy Koehler for help with the figures; Fotis Kafatos for use of facilities; Philip Sharp, Al Bothwell, Shirley Tilghman, Doug Hanahan, and Richard Firtel for discussions. W.G. is an American Cancer Society Professor of Molecular Biology. W.L.C. is an Established Investigator of the American Diabetes Association. This work was supported by National Institutes of Health Grants AM 21240 and GM 09541-17 to W.G. and AM 15398 to W.L.C.

1. Chan, S. J., Keim, P. & Steiner, D. F. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 1964–1968.
2. Chan, S. J. & Steiner, D. F. (1977) *Trends Biochem. Sci.* **2,** 254–256.
3. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67,** 835–851.
4. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67,** 852–862.
5. Steiner, D. F., Kemmler, W., Tager, H. S. & Peterson, J. D. (1974) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **33,** 2105–2115.
6. Smith, L. F. (1966) *Am. J. Med.* **40,** 662–666.
7. Clark, J. L. & Steiner, D. F. (1969) *Proc. Natl. Acad. Sci. USA* **62,** 278–285.
8. Markussen, J. & Sundby, F. (1972) *Eur. J. Biochem.* **25,** 153–162.
9. Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W. J. & Goodman, H. M. (1977) *Science* **196,** 1313–1319.
10. Chick, W. L., Warren, S., Chute, R. N., Like, A. A., Lauris, V. & Kitchen, K. C. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 628–632.
11. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crossa, J. H. & Falkow, S. (1977) *Gene* **2,** 95–113.
12. Sutcliffe, J. G. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3737–3741.
13. Ambler, R. P. & Scott, G. K. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3732–3736.
14. Boyer, H. W. & Rouland-Dussoix, D. (1969) *J. Mol. Biol.* **41,** 459–472.
15. Curtiss, R., III, Pereira, D. A., Hsu, J. C., Hull, S. C., Clarke, J. E., Maturin, L. J., Sr., Goldschmidt, R., Moody, R., Inoue, M. & Alexander, L. (1977) in *Recombinant Molecules: Impact on Science and Society. Proceedings of the 10th Miles International Symposium,* eds. Beers, R. F., Jr., & Bassett, E. G. (Raven, New York), pp. 45–56.
16. Clewell, D. B. (1972) *J. Bacteriol.* **110,** 667–676.
17. Palmiter, R. (1974) *Biochemistry* **13,** 3603–3615.
18. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
19. Efstratiadis, A., Kafatos, F. C., Maxam, A. M. & Maniatis, T. (1976) *Cell* **7,** 279–288.
20. Roychoudhury, R., Jay, E. & Wu, R. (1976) *Nucleic Acid Res.* **3,** 101–116.
21. Gilbert, W. & Maxam, A. M. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 3581–3584.
22. Enea, V., Vovis, G. F. & Zinder, N. D. (1975) *J. Mol. Biol.* **96,** 495–509.
23. Miller, J. M. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York), pp. 431–435.
24. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 3961–3965.
25. Paterson, B. M., Roberts, B. E. & Kuff, E. L. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4370–4374.
26. Roberts, B. E. & Paterson, B. M. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 2330–2334.
27. Lomedico, P. T. & Saunders, G. F. (1976) *Nucleic Acids Res.* **3,** 381–391.
28. Broome, S. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2746–2749.
29. Makula, D. R., Vichnuk, D., Wright, P. H., Sussman, K. E. & Yu, P. L. (1969) *Diabetes* **18,** 660–689.
30. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
31. Boyer, H. W., Betlach, M., Bolivar, F., Rodriguez, R. L., Heyneker, H. L., Shine, J. & Goodman, H. M. (1977) in *Recombinant Molecules: Impact on Science and Society. Proceedings of the 10th Miles International Symposium,* eds. Beers, R. F., Jr. & Bassett, E. G. (Raven, New York), pp. 9–20.
32. Casey, J. & Davidson, N. (1977) *Nucleic Acids Res.* **4,** 1539–1552.
33. Humbel, R. E., Bosshard, H. R. & Zahn, H. (1972) in *Handbook of Physiology, Section 7 (Endocrinology),* eds. Steiner, D. F. & Freinkel, N., (American Physiological Society, Washington, DC), Vol. 1, pp. 111–132.
34. Neu, H. C. & Heppel, L. A. (1965) *J. Biol. Chem.* **240,** 3685–3692.
35. Tilghman, S. M., Tiemeier, D. C., Seidman, J. G., Peterlin, B. M., Sullivan, M., Maizel, J. V. & Leder, P. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 725–729.
36. Jeffreys, A. J. & Flavell, R. A. (1977) *Cell* **12,** 1097–1108.
37. Breathnach, R., Mandel, J. L. & Chambon, P. (1977) *Nature* **270,** 314–319.
38. Gilbert, W. (1978) *Nature* **271,** 501.

TAB L

Nature Vol. 275 12 October 1978

5. Bilezikian, S. B., Nossel, H. L., Butler, V. P. Jr & Canfield, R. E. *J. clin. Invest.* **56**, 438–445 (1975).
6. Herrig, K. H., Rutnoff, O. D. & Shainoff, J. R. *J. Lab. clin. Med.* **76**, 451–465 (1970).
7. Shainoff, J. R. & Dardik, B. N. *Thromb. Haemostasis* **38**, 193 (1977). Abstr.
8. Slomback, B. & Laurent, T. C. *Ark. Kemi* **12**, 137–146 (1958).
9. Steiner, R. F. & Laki, K. *Archs Biochem. Biophys.* **34**, 24–37 (1951).
10. Laurent, T. C. & Blombäck, B. *Acta chem. scand.* **12**, 1875–1877 (1958).
11. Blombäck, B. & Blombäck, M. *Ann. N.Y. Acad. Sci.* **202**, 77–97 (1972).
12. Edgar, W. & Prentice, C. R. M. 1st *Florence Conf. Haemostasis Thrombosis.* (eds Neri Serneri, G. G., Gensini, G. F. & Abbate, R. A.) 45 (1977) (Abstract).
13. Kudryk, B. J., Collen, D., Woods, K. R. & Blombäck, B. *J. biol. Chem.* **249**, 3322–3325 (1974).
14. York, L. J. & Blombäck, B. *Thromb. Res.* **8**, 607–618 (1976).
15. Kudryk, B., Blombäck, B. & Blombäck, M. *Thromb. Res.* **9**, 25–36 (1976).
16. Shen, L. L., Hermans, J., McDonagh, J. & McDonagh, R. P. *Am. J. Physiol.* **232**, H629–H633 (1977).
17. Hogg, D. H. & Blombäck, B. *5th Cong. Int. Soc. Thromb. Haemostasis*, Paris, Abstr. 26 (1975).
18. Blombäck, B., Blombäck, M., Hessel, B. & Iwanaga, S. *Nature* **215**, 1445–1448 (1967).
19. Ferry, G. D. & Morrison, P. R. *J. Am. chem. Soc.* **69**, 388–400 (1947).
20. Seegers, W. H. & Smith, H. P. *Am. J. Physiol.* **137**, 348–354 (1942).
21. Godal, H. C. *Scand. J. clin. Lab. Invest.* **12**, Suppl. 53, 3–20 (1960).
22. Blombäck, B., Blombäck, M., Laurent, T. C. & Peterli, H. *Biochim. biophys. Acta* **127**, 560–562 (1966).
23. Blombäck, B., Gröndahl, N. J., Hessel, B., Iwanaga, S. & Wallen, P. *J. biol. Chem.* **248**, 5806–5820 (1973).
24. Inada, Y., Hessel, B. & Blombäck, B. *Biochim. biophys. Acta* **532**, 161–170 (1978).
25. Blombäck, M., Blombäck, B., Mammen, E. F. & Prasad, A. S. *Nature* **218**, 134–137 (1968).
26. Mammen, E. F., Prasad, A. S., Barnhart, M. I. & Au, C. C. *J. clin. Invest.* **48**, 235–249 (1969).
27. Blombäck, B. & Blombäck, M. *Nouv. Revue fr. Hémat.* **10**, 671–678 (1970).
28. Nossel, H. L. *et al. J. clin. Invest.* **58**, 1136–1144 (1976).
29. Blombäck, B., Hessel, B. & Hogg, D. *Thromb. Res.* **8**, 639–658 (1976).
30. Marguerie, G., Chagniel, G. & Sussillon, M. *Biochim. biophys. Acta* **490**, 94–103 (1977).
31. Ly, B. & Godal, H. C. *Haemostasis* **1**, 204–209 (1972).
32. Kwaan, H. C., Barlow, G. H. & Suwanwela, N. *Thromb. Res.* **2**, 123–136 (1973).
33. Gaffney, P. J. & Brasher, M. *Nature* **251**, 53–54 (1974).
34. McKillap, C., Edgar, W., Focker, C. D. & Prentice, C. R. M. *Thromb. Res.* **7**, 361–372 (1975).
35. Hogg, D. & Blombäck, B. *Thromb. Res.* **5**, 685–693 (1974).

Synthesis of an ovalbumin-like protein by *Escherichia coli* K12 harbouring a recombinant plasmid

Odile Mercereau-Puijalon, André Royal, Brigitte Cami, Axel Garapin, Andrée Krust*, Frank Gannon* & Philippe Kourilsky

Groupe de Biologie Moléculaire du gène, ERA CNRS 201, Unité de Génie Génétique, Institut Pasteur, 28, Rue du Dr. Roux, 75724 Paris Cedex 15

A cloned DNA transcript of ovalbumin mRNA was cut a few nucleotides away from the initiator codon, and fused in phase to the beginning of the Escherichia coli *β-galactosidase gene. The hybrid gene has been cloned in* E. coli *where it produces large amounts of an ovalbumin-like protein.*

RECOMBINANT DNA technology now permits the introduction and cloning, in *Escherichia coli* K12, of genes of various origins. However, the expression of foreign genes in this bacterial environment is still poorly documented. Several genes isolated from the lower eukaryotes (yeast, *Neurospora crassa*, *Drosophila*) have been shown to be spontaneously expressed in *E. coli*[1–6]. There has been no report of genes cloned from the higher eukaryotes that are spontaneously transcribed or translated. Transcription can be efficiently enforced by the integration of gene sequences within a bacterial or a phage gene[7,8]. It seems that the best way to promote translation into a polypeptide product is by the construction of hybrid genes in which the eukaryotic sequence is fused in phase to a bacterial gene. This approach has been used to construct *E. coli* strains capable of producing the small hormone, somatostatin, in the form of a hybrid protein[9]. The gene coding for the hormone, which is 14 amino acids long, was synthesised chemically, fused to the end of the *E. coli lac Z* gene coding for β-galactosidase and cloned in a multicopy plasmid. The β-galactosidase moiety (about 1,000 amino acids) was elegantly connected to the somatostatin moiety through a methionine residue. As somatostatin contains no methionine, and as the latter is labile in certain conditions, the hormone could be extracted from the hybrid polypeptide by treatment with cyanogen bromide[9].

Most polypeptides, however, contain methionine residues, so that the above method is not of potential general use unless other cleavable protein 'linkers' are made available. Many possible applications of genetic engineering, on the other hand, rely on the synthesis of polypeptides larger than somatostatin, coded by gene sequences too long to be chemically synthesised at present. The recent discovery that at least some eukaryotic genes are split[10–20] makes it unlikely that such genes, isolated from the genome, are easily amenable to expression in *E. coli*. Rather, the most general approach seems to involve (1) enzymatic synthesis of the desired gene sequence from mRNA; (2) fusion of this cDNA sequence to a bacterial gene in such a way that the eukaryotic sequence can be read in the proper phase; (3) cloning in *E. coli*.

We have applied this protocol to the chicken ovalbumin gene sequence and we report here the construction of bacterial strains which synthesise large amounts of an ovalbumin-like protein. This result demonstrates the practicability of the approach and opens broader prospects to genetic engineering methodology.

The ovalbumin system

Ovalbumin is the major egg white protein. It is a polypeptide chain of 386 amino acids, with a single intra-chain disulphide bond, and undergoes a few post-translational modifications such as cleavage of the first methionine residue, *N*-acetylation, glycosylation and phosphorylation of certain residues. There are several known genetic variants[21–26].

Ovalbumin mRNA (ov mRNA) is predominant in the chick oviduct, which facilitates its isolation in relatively pure form. Ov mRNA is 1,859 nucleotides long. Its complete sequence has been determined by McReynolds *et al.*[26,27], and consists of a 64-nucleotide long 5' non-coding region, followed by a 1,158-long coding region, and a 637-long 3' non-coding region (Fig. 1).

The ovalbumin gene is split in the chicken genome[12,16,20]. We have recently isolated, by molecular cloning, the various genomic DNA fragments which code for cytoplasmic ovalbumin mRNA[16,17]. Analysis of the cloned fragments has revealed the presence of at least six intervening sequences, all located within translated regions of ov mRNA. As the structure of the

* Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS et U. 44 de l'INSERM, Faculté de Médicine, Strasbourg, France.

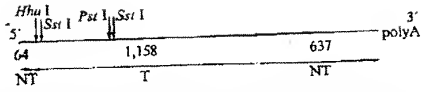


Fig. 1 Map of ovalbumin mRNA. Schematic representation of ov mRNA according to McReynolds *et al.*[26], showing the translated (T) and non-translated (NT) regions, the AUG start signal (nucleotides 65–67) for protein synthesis, and the position of the *Hha*I, *Pst*I and *Sst*I sites in the corresponding DNA sequence[26,31].

ovalbumin gene was likely to prevent its expression in *E. coli*, we turned our attention towards the DNA transcript of ov mRNA.

Humphries *et al.*[28], using a protocol devised for the cloning of globin mRNA transcripts[29,30], first isolated a plasmid, pCRI ov2-1, which carries about 1,730 base pairs of DNA transcribed from ov mRNA. Evidence has been presented that a faithful cloned DNA copy is obtained[28]. As ov mRNA is 1,859 nucleotides long, pCRI ov2-1 does not harbour all the ov mRNA sequence. Restriction mapping, DNA sequencing and electron microscopy analysis[16,17,28,31], however, indicate that the missing parts originate from untranslated regions of the mRNA.

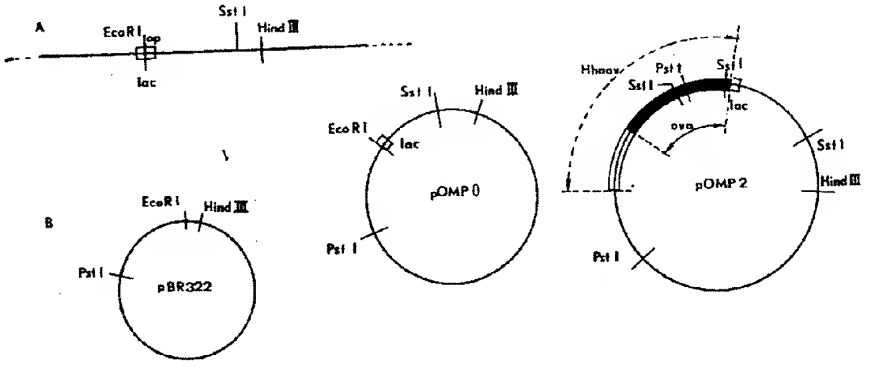
As pCRIov2-1 carries all the sequence coding for ovalbumin, we first asked whether the gene is expressed in C600$r_k^-m_k^+$ (pCRIov2-1). Transcription of the ovalbumin sequence from the correct strand could be demonstrated. However, the strain produced no ovalbumin-like protein, as indicated by radioimmunoassays (see below) which were negative, even in conditions where less than one molecule per cell would have been detected (O.M.-P., unpublished). We tentatively concluded that the ovalbumin sequence in pCRIov2-1 has been accidentally fused to an uuknown plasmid gene, probably out of the proper reading frame.

Construction of recombinant *lac*-ov plasmids

We then decided to fuse the ovalbumin gene sequence to the beginning of the *E. coli lac Z* gene, which codes for β-galactosidase. Fuller[32] has generated a small DNA fragment of 205 base pairs, flanked by two *Eco*RI sites, which carries, in addition to the L8UV5 *lac* operator promoter region, the information for the first eight amino acids of β-galactosidase. (The L8UV5 mutation makes the system insensitive to catabolite repression.) This fragment was engineered by Charnay *et al.*[33] into a phage, λcI857p*lac*5ΔZL8UV5, the DNA of which has a single *Eco*RI site located after the eighth amino acid of β-galactosidase in the *lac* sequence. We used this DNA to generate by digestion with *Eco*RI and *Hind*III a DNA fragment of 2.85 kilobases which carries the beginning of the *lac* sequence. This fragment was recombined *in vitro* with pBR322 DNA[34] digested by the same two enzymes, as outlined in Fig. 2, and cloned in *E. coli*. The resulting plasmid, called pOMP0, was the vector used for further recombination experiments. pOMP0 is about 7.1 kilobases in size. It makes the recipient host cell resistant to ampicillin (but not to tetracycline) and *lac* constitutive (because *lac* repressor is overtitrated by the operator carried by the multicopy plasmid), and has a single *Eco*RI site located near the beginning of the *lac Z* gene (Fig. 2).

Most of the ovalbumin sequence carried by plasmid pCRIov2-1 can be isolated in a restriction fragment of 2,430 base pairs, *Hha*ov, obtained by digestion of plasmid DNA with *Hha*I[12,28]. The analysis of cloned genomic parts of the ovalbumin gene[17] has demonstrated the existence, near the 5′ end, of a site sensitive to *Hha*I, located very close (about 30 base pairs) to an *Sst*I site in a region present in pCRIov2-1 (ref. 31). When the nucleotide sequence of this region of ov mRNA established by McReynolds *et al.*[26] became available, it seemed that there was no *Hha*I site (GCGC) in the sequence. It occurred to us, however, that 35 nucleotides ahead of the *Sst*I site closest to the 5′ end (nucleotides 116–120), there was a sequence, GGUGCA (nucleotides 77–82), which could be changed to GGCGCA without modifying the amino acid sequence (GGU and GGC both code for glycine). The assumption could be made that McReynolds *et al.*[26] and Humphries *et al.*[28] had worked on different genetic variants and that the *Hha*I site which we observed would be located at nucleotides 78–81 of ov mRNA, overlapping two triplets, GGC and GCA, which code for the fifth and sixth amino acids of ovalbumin. Accordingly, the fragment extracted from pCRIov2-1 should be cut at the very beginning of the mRNA sequence coding for amino acids.

Fig. 2 Construction of plasmids pOMP0 and pOMP2. λcI857 *plac*5 L8UV5ΔZ was constructed by Charnay *et al.*[33] by first inserting the 205-base pair *lac* sequence (described in ref. 32) into the *Eco*RI site located at the end of the *lac Z* gene of a derivative of λcI857p*lac*5L8UV5. Intramolecular recombination subsequently yielded a phage with only one *Eco*RI site between the end of the *lac Z* gene and the 205-base pair *lac* sequence as shown in *a*. 500 ng of λcI857p*lac*5L8UV5ΔZ DNA (shown as *a*, where only the region carrying the *lac* sequence is depicted) and 50 ng of pBR322 DNA (shown as *b*) were digested by *Eco*RI and *Hind*III, and ligated by T4 DNA ligase in a final volume of 15 μl as usual[16,35]. Aliquots were used to transform $CaCl_2$-treated strain 478 (C600 $r_k^-m_k^-$recBC$^-$*lacy*$^-$) and transformants were selected on tryptone plates containing ampicillin (20 μg ml^{-1}) and 5-bromo-4-chloro-indolylgalactoside (X-gal, 40 μg ml^{-1}), on which strains carrying the recombinant plasmid give blue colonies. Purified pOMP0 was next used to construct recombinant plasmids harbouring the ovalbumin sequence (see legend to Fig. 3). The map of one representative plasmid, pOMP2, was established by restriction analysis using the indicated sites (see Fig. 4). pOMP0 is about 7.1 kilobases long, has single sites for *Eco*RI, *Hind*III, *Sst*I and *Pst*I. All experiments were carried out with ultrapure hyperhydor water (Institut Pasteur).



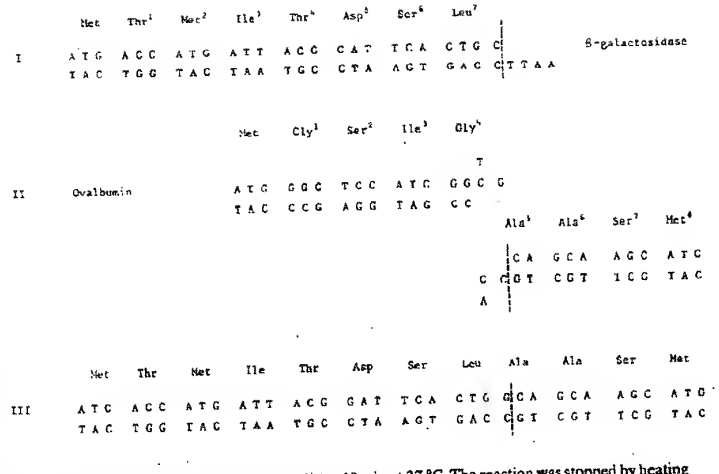


Fig. 3 Fusion of the ovalbumin sequence to the beginning of the *lac Z* gene. The DNA sequence of the *Eco*RI-restricted *lac Z* gene fragment is shown in I with the corresponding amino acid sequence[39]. The first eight amino acids of ovalbumin and their corresponding DNA sequence[26] are shown in II, with one change at position 79 (C:G instead of T:A written above the line) which creates a site for *Hha*I restriction endonuclease[31]. The latter splits the DNA as indicated. Elimination of the single-stranded ends and blunt-end ligation produce the fused gene sequence shown in III, with the corresponding amino acid sequence. Single-stranded ends were eliminated in the case of pOMP0 by digestion with S_1 nuclease followed by a treatment with DNA polymerase I (to repair possible mistakes made by S_1). *Hha*ov protruding ends were digested either by S_1 nuclease or S_1 nuclease followed by DNA polymerase I, or by DNA polymerase I alone making use of the 3′ exonuclase activity of this enzyme. 1 μg of purified[28] pOMP0 DNA was digested by *Eco*RI (ref. 43) for 40 min at 37 °C. The reaction was stopped by heating for 10 min at 65 °C. The mixture was adjusted to 3 mM $ZnSO_4$, 30 mM Na acetate, 300 mM NaCl, and 20,000 units of purified S_1 endonuclease[28] were added. After 15 min at 37 °C, the reaction was stopped by heating for 10 min at 65 °C, and the DNA was dialysed against 10 mM Tris-HCl, *p*H 7.5, 5 mM EDTA (step I). Part of this preparation (330 ng) was treated, in the presence of all 4 dNTPs, with 0.2 units of DNA polymerase I (ref. 28) for 15 min at 15 °C and then heated for 10 min at 65 °C (step II). Part of the DNA was then treated by bacterial alkaline phosphatase (Worthington, heated for 5 min at 80 °C before use) for 15 min at 65 °C (ref. 51). The reaction was terminated by extraction with a mixture of phenol and chloroform, and the DNA was extensively dialysed against 10 mM Tris-HCl, *p*H 7.5, 5 mM EDTA (step III). *Hha*ov (ref. 28) fragment was treated either with: *a*, S_1 exonuclease (500 ng of *Hha*ov were digested by 5,000 units of S_1 in the conditions described for pOMP0) or, *b*, DNA polymerase I (280 ng of *Hha*ov were treated with 0.2 unit of enzyme as described for pOMP0) or, *c*, with S_1 exonuclease and DNA polymerase I (150 ng of the DNA preparation in *a* being treated with 0.2 unit of DNA polymerase I). 20 ng of pOMP0 DNA at steps I, II or III was ligated with 7 ng of *Hha*ov treated either as in *a*, *b* or *c* in a final volume of 10 μl with 0.12 units of T4 DNA ligase (Biolabs) for 15 h at 12 °C (refs 9, 51). The ligated mixture was diluted to 200 μl with 0.1 M Tris-HCl, *p*H 7.0, and used to transform $CaCl_2$-treated *E. coli* 1398 (ref. 35). Transformants were selected on tryptone plates containing ampicillin (20 μg ml^{-1}). Colonies were directly screened by *in situ* hybridisation with ^{32}P-labelled *Hha*ov for the presence of ovalbumin sequence[35]. The positive clones were tested for *lac* constitutivity on tryptone X-gal (40 μg ml^{-1}) plates. DNA from the *lac*$^+$, ov$^+$, ap^R clones was then used to transform C600$r_k^-m_k^+$ and 1442. The plasmids described here arose from the following combinations: pOMP1, 10 and 11 = step II pOMP0 + *Hha*ov *b*; pOMP2 and 7 = step II pOMP0 + *Hha*ov *a*; pOMP4 = step III pOMP0 + *Hha*ov *b*; pOMP5 and 9 = step III pOMP0 + *Hha*ov *c*.

Sequencing of pCRIov2-1 DNA[31] later demonstrated that these assumptions were correct (Fig. 3).

The way in which the ovalbumin sequence in *Hha*ov was fused in phase to the beginning of the *Z* gene is depicted in Fig. 3. In brief, after elimination of the single-stranded ends of the *Hha*ov fragment and of *Eco*RI-treated pOMP0 DNA by treatment with S_1 nuclease and/or DNA polymerase I, blunt-end ligation with an excess of T4 DNA ligase should yield the structure shown in Fig. 3, in those strains where integration of the *Hha*ov fragment has occurred in the proper orientation. The hybrid gene is then expected to produce an ovalbumin-like protein of 389 amino acids in which the first five amino acids of ovalbumin are substituted by the first eight amino acids of β-galactosidase (including the first Met residue).

Recombinant molecules were introduced by transformation into *E. coli* 1398 (ref. 35), a lysogenic host adequate for the *in situ* colony hybridisation method which we have recently devised[35]. Ampicillin-resistant colonies were then screened with ^{32}P-labelled *Hha*ov, a probe which reacts only with recombinant plasmids harbouring an ovalbumin sequence. We thus obtained eight independent strains showing *lac* constitutivity (blue colonies on X-gal plates, see legend to Fig. 3). The isolated plasmids (pOMP1, 2, 4, 5, 7, 9, 10, 11) were then transferred into a non-lysogenic host (C600$r_k^-m_k^+$).

Characterisation of the recombinant plasmids

The orientation of the integrated ovalbumin sequence was determined as follows: recombinant plasmids are cut by *Pst*I in two sites, one in the ampicillin gene of pBR322 (ref. 34), the other in the ovalbumin sequence[12]. If the latter is properly orientated, digestion of plasmid DNA by *Pst*I should yield two DNA fragments of about 6.9 and 2.7 kilobases. If it is not, digestion by *Pst*I should produce fragments of about 8.5 and 1.1 kilobases. The results shown in Fig. 4 indicate that the isolated plasmids do fall into these two expected classes: in plasmids pOMP1, 2 and 4, the ovalbumin sequence is in the proper orientation to be transcribed from the *lac* promoter, whereas in plasmids pOMP5, 7, 10 and 11, it is in the opposite direction.

To determine whether some of these strains produce an ovalbumin-like protein, we used a radioimmunoassay, in which iodinated ovalbumin is mixed to bacterial extracts and then allowed to react with anti-ovalbumin specific antibodies. The details of the method are described in the legend to Fig. 5. The results were perfectly correlated with the structure of the recombinant plasmids: plasmids pOMP1, pOMP2 and pOMP4 caused very strong competition, whereas the others (including pOMP0) did not. Plasmid pOMP2 was arbitrarily selected for further studies.

Production of an ovalbumin-like protein by pOMP2

Further radioimmunoassays were carried out to quantitate the amount of ovalbumin-like product made in *E. coli*. In Fig. 5, competition by pure, unlabelled ovalbumin (Fig. 5*b*) is compared to competition by extracts of an *E. coli* strain carrying pOMP2 (Fig. 5*a*). The competition by the latter is only 80% complete, which suggests that one (or several) antigenic site is missing in the protein manufactured by bacteria. Furthermore, the competition curve may be regarded as being multiphasic. This suggests either that some of the antigenic sites made in *E. coli* are more abundant than others, or that some have a lower affinity than those of native ovalbumin. Because the size of the

ovalbumin-like product is approximately that expected (Fig. 6 and below), we favour the second interpretation. As we noted above, ovalbumin, in the oviduct, undergoes some post-translational modifications which may play an important part in the structure of antigenic determinants. As it is unlikely that such modifications take place in *E. coli*, the ovalbumin-like product (which in any case differs from native ovalbumin by a few amino acids in the N-terminal end) may be, for some of the antigenic sites, a less efficient competitor than ovalbumin. On this assumption, quantitation based on the first part of the multiphasic curve indicates that 2.5×10^6 cells may produce the equivalent of about 15 ng of ovalbumin, that is, about 90,000 molecules of ovalbumin-like protein per cell. If the whole curve was considered as monophasic the estimate would be of 45,000 molecules per cell (5,000–45,000 in a variety of independent experiments, 30,000 on average).

Fig. 4 Restriction analysis of plasmids carrying the ovalbumin sequence. Horizontal agarose (0.5% Sigma type II) gels (25 cm long) were run at 5 V cm^{-1} as in ref. 40. Plasmid DNAs were purified from cleared lysates by phenol extraction and ethanol precipitation, subjected to digestion by *Pst*I endonuclease (Boehringer)[41,42], and run in parallel with size markers (lanes 5 and 10: λcI857*Sam*7 DNA digested by *Eco*RI and *Bam*H1[43,44], lane 11: λcI857S7 DNA digested by *Eco*RI (ref. 43), lane 12: pBR322 DNA digested by *Hin*dIII (ref. 34). The size of fragments was estimated on a semilogarithmic plot relating distance to size (in kilobase pairs). Plasmids pOMP1, 2 and 4 (lanes 1, 2 and 3, respectively) show two fragments of estimated size 6.9 and 2.8 kilobases. Plasmids pOMP5, 7, 10 and 11 (lanes 4, 6, 8 and 9, respectively) show two fragments of about 8.65 and 0.7 kilobases. Plasmid pOMP9 (lane 7) does not yield the expected fragments. The total plasmid size (except for pOMP9) can thus be estimated to range over 9.35–9.8 kilobases. The theoretical size is 9.63 kilobases (using estimates of 4.35 kilobases for pBR322; 2.43 kilobases for *Hhaov*; 2.85 kilobases for the λ*lac Eco*RI-*Hin*dIII fragment (as determined by electron microscopy). The respective orientation of the *lac* and ovalbumin sequences which can be deduced from these data was confirmed, in the case of pOMP2, by a variety of other experiments including double digestion by *Pst*I and *Hin*dIII, which yields three fragments of about 3.45, 3.1 and 2.97 kilobases, and digestion by *Sst*I (data not shown) (see Fig. 2). Lane 13, plasmid pOMP2 digested by *Hin*dIII (ref. 34).

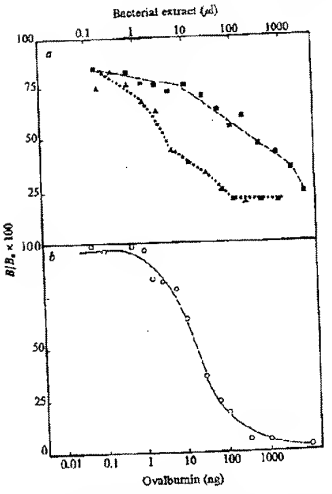


Fig. 5 Detection of ovalbumin-like protein in bacterial extracts. Ovalbumin (Sigma, 99% pure) (30 μg in 20 μl) was labelled with 1 mCi of Bolton and Hunter reagent (Amersham; >1,375 Ci mmol^{-1}), according to the procedure of ref. 45, except that the reaction was stopped after 18 h at 0 °C. Radioactive protein was then applied to a Sephadex-G75 column equilibrated with 0.05 M phosphate buffer, pH 7.5, 0.25% gelatin, and eluted with the same buffer. The iodinated product (initial specific activity ~16.5 μCi μg^{-1}) was stored at 4 °C. Antibodies against ovalbumin were raised in rabbits by injection of 1 mg of ovalbumin (purified by preparative gel electrophoresis) emulsified in Freund's complete adjuvant. A second injection was made 3 weeks later, and bleedings were made 2 weeks after. Antibodies were purified by $(NH_4)_2SO_4$ precipitation and chromatography on CM-DEAE cellulose and Biogel P11 columns[47]. Radioimmunoassays were carried out in a final volume of 300 μl in PBSA (phosphate buffer 0.1 M, pH 7.5, NaCl 0.15 M, bovine serum albumin 5 mg ml^{-1}). To the desired dilution of bacterial extract (or of unlabelled ovalbumin) in PBSA, were added 20 μl of mixture containing 15 μl of nonspecific rabbit antiserum as a carrier and 10,000 c.p.m. iodinated ovalbumin. This was followed by addition of 50 μl of a dilution of the specific antibodies able to precipitate 50% of the iodinated protein. After an 18-h incubation at 4 °C, antibody-bound ovalbumin was separated from free ovalbumin by addition of Na_2SO_4 18% (w/v) final. The precipitate was centrifuged for 5 min at room temperature in a microfuge; supernatants were pipetted and pellets and supernatants counted in a gamma counter. Control experiments with iodinated antibodies demonstrated that precipitation by Na_2SO_4 was essentially complete. Precipitation of iodinated ovalbumin in the absence of specific antibodies did not exceed 5–8% of input. Strain 1442 (BMH 7156 Δ(prulac) F'lac^{Oi}Z u118 was transformed by pOMP2, and grown on glucose minimum plates (to prevent loss of the episome ensuring the overproduction of *lac* repressor). Liquid cultures of 1442 (pOMP2) and C600r$_k$m$_k^+$(pOMP2) in L broth with 5 μg ml^{-1} ampicillin were inoculated from plates at about 10^8 cells per ml and grown to about 1.2×10^9 bacteria per ml in the presence or absence of IPTG (5 or 10 mM). In some experiments cultures were directly frozen and sonicated in the presence of protease Inhibitor, PTSF[46] *p*-toluyl-sulphonyl-fluoride, Ega-Chemie, 30 μg ml^{-1}). In the experiment shown in *a*, bacteria were concentrated 20-fold in PBSA containing PTSF, frozen, sonicated (2 × 20 s in the cold) and diluted serially in the same buffer mixed with a constant amount of extract of the parent strain 1441. Radinimmunoassays were carried out in duplicate. B/B_0 is the ratio between the c.p.m. precipitated in the presence and the absence of bacterial extracts or unlabelled ovalbumin (nonspecifically precipitated c.p.m. being subtracted each time). *a*, Extracts of 1442 (pOMP2) with (▲) and without IPTG (■). The abscissa has been recalculated as μl of nonconcentrated extract (1 μl = 1.2×10^6 bacteria). *b*, Standard curve with ovalbumin (Sigma, 99% pure). 50% competition corresponds to 15 ng of pure ovalbumin.

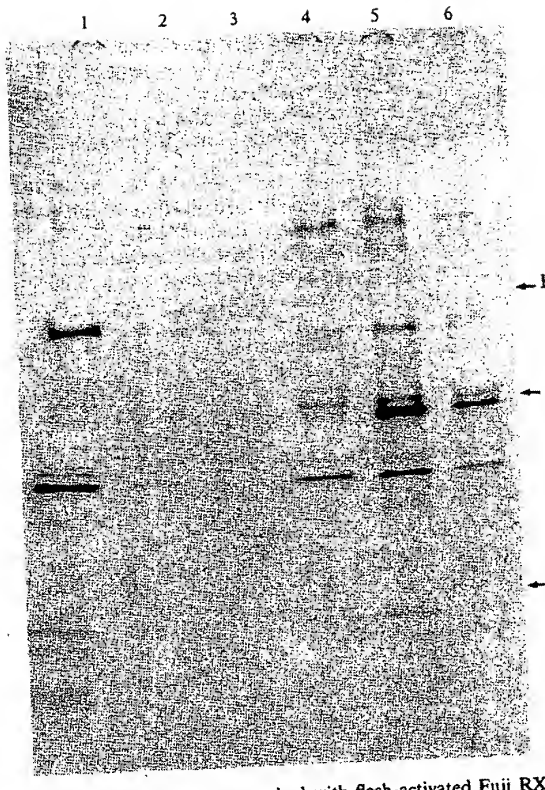


Fig. 6 Size of the ovalbumin-like product in immunoprecipitated bacterial extracts. 5-ml cultures of 1442(pOMP0) (lanes 1 and 2) and 1442(pOMP2) (lanes 3–6) were inoculated at 37 °C in 63B1 glycerol medium supplemented with 5 μg ml^{-1} ampicillin and eventually (lanes 5 and 6) with 1 mM IPTG, grown to an A_{600} of 0.5, labelled with ^{35}S-methionine (Amersham, 50 μCi ml^{-1}, 800 Ci mMol^{-1}) for 30 min and poured over an equal volume of a crushed frozen solution containing 10 mM NaN_3, 200 μg ml^{-1} chloramphenicol and 30 μg ml^{-1} PTSF. The cells were then centrifuged (5,000g, 10 min at 4 °C) and frozen at −20 °C. The pellet was resuspended on 0.5 ml of 10 mM Na phosphate buffer, *p*H 6.8, and sonicated (2×20 s in ice). Immunoprecipitates were prepared as described by Rhoads *et al.*[47] with minor modifications. To 10^7 trichloroacetic acid-precipitable c.p.m. were added either 5 μg of unlabelled carrier ovalbumin and 50 μg of bovine serum albumin (BSA) (lanes 3 and 5), or 50 μg of competing unlabelled ovalbumin and 5 μg of BSA (lanes 2, 4 and 6). Samples were adjusted to 1.4% Triton X-100, 1.4% sodium deoxycholate, 10 mM Na phosphate buffer, *p*H 7.5, and 0.15 M NaCl, to which 20 μl of a nonspecific rabbit serum were added. The mixture (210 μl) was incubated for 30 min at room temperature, layered over 200 μl of 1 M sucrose, 1% Triton X-100, 1% deoxycholate, 10 mM Na phosphate buffer, *p*H 7.5, and 0.15 NaCl, and centrifuged (5 min at 4 °C in a microfuge). The upper phase was carefully pipetted and 20 μl of rabbit anti-ovalbumin antibodies, enough to precipitate 7.5 μg of ovalbumin, were added. After 60 min at room temperature, the mixture was layered over sucrose as described above, centrifuged and frozen in liquid N_2. A 3-mm long tip was cut from the bottom of the tube and the immunoprecipitate was resuspended in 20 μl of 10% glycerol, 2% sodium dodecyl sulphate, 62.5 mM Tris-HCl, *p*H 6.8, 0.7 M β-mercaptoethanol and bromophenol blue, heated for 2 min at 100 °C and loaded on a sodium dodecyl sulphate polyacrylamide gel. The resolving gel was a 6–15% linear gradient of acrylamide[48]. The electrophoresis was carried out with the apparatus described by Studier[49] at 4 °C for 12 h at 5 V. After the run, the gel was fixed and stained with Coomassie brilliant blue R-250 (0,2% (w/v) in methanol:water:acetic acid, 50:50:7 v/v), impregnated with diphenyloxazole, dried and fluorographed with flash-activated Fuji RX films according to the standard technique[50]. Exposure was for 24 h at −70 °C. Arrows indicate the position of unlabelled size markers added to each sample: bovine serum albumin (BSA, 68,000 daltons), ovalbumin (OV) and trypsin inhibitor (TI, 21,500 daltons). The immunoprecipitates are contaminated by several bacterial proteins. This is probably due to a weak anti-*E. coli* activity of the rabbit antibodies used. Lanes 1 and 2, 1442(pOMP0); lanes 3 and 4, 1442(pOMP2) without addition of IPTG; lanes 5 and 6; 1442 (pOMP2) grown in 1 mM IPTG. In lanes 2, 4 and 6, excess unlabelled ovalbumin was added as a competitor.

We could not clearly demonstrate, in C600 (pOMP2), that the synthesis of the ovalbumin-like product is under the control of the *lac* repressor (which is, at least partially, overtitrated by the operators). Because of this, we introduced pOMP2 into a strain (1442) which overproduces *lac* repressor. Synthesis was then stimulated by the *lac* inducer IPTG about 50-fold (Fig. 5*a*).

The size of the ovalbumin-like product was estimated as follows: extracts of cells harbouring pOMP2 and labelled *in vivo* with ^{35}S-methionine were precipitated by anti-ovalbumin antibodies, and the dissociated precipitates were loaded on polyacrylamide gels in denaturing conditions as explained in the legend to Fig. 6. The fluorograms in Fig. 6 show a predominant band corresponding to a protein which either co-migrates with or migrates slightly more rapidly than ovalbumin, depending on electrophoresis conditions. In strain 1442 (pOMP2), synthesis of this protein depends on the addition of IPTG. Addition of excess unlabelled ovalbumin in the immunoprecipitation eliminates most of the radioactive band, which indicates that it corresponds to an ovalbumin-like protein. In total labelled extracts not subjected to immunoprecipitation, the band could sometimes be directly observed. In summary, the ovalbumin-like product has a size close to that of ovalbumin.

Synthesis of foreign polypeptides by *E. coli*

We have constructed *E. coli* strains capable of synthesising an ovalbumin-like product. The available evidence indicates that: (1) synthesis of this product is under *lac* repressor control; (2) the product competes efficiently with iodinated ovalbumin in a radioimmunoassay; (3) the size of the product is very similar to that of ovalbumin.

Together, these data strongly suggest that our *E. coli* strains synthesise an ovalbumin-like product, the composition of which is probably very close to that expected, that is, eight amino acids of β-galactosidase and 381 amino acids of ovalbumin (out of the 386 of native ovalbumin). The exact structure of the hybrid protein remains to be investigated.

Bacteria harbouring the recombinant plasmids synthesise an average of 30,000 or 60,000, and up to 45,000 or 90,000 molecules per cell (depending on the interpretation of the radioimmunoassay competition curves). This corresponds to about 0.5–1% of the total protein mass of an *E. coli* cell. These figures are relatively close to the theoretical yield which can be estimated (from the number of β-galactosidase monomers synthesised in cells harbouring a multicopy plasmid with a functional *Z* gene) to be of the order of $1–2\times10^5$ molecules per cell.

The production of high quantities of an ovalbumin-like protein by *E. coli* strains raises several interesting points. First, the strains are perfectly stable and show no tendency to lose the recombinant plasmid. Second, ovalbumin is a secreted protein; in preliminary experiments, we find no trace of ovalbumin-like product in the supernatant of C600 (pOMP2) cultures. Third, the ovalbumin-like protein made by *E. coli* must be relatively stable in the bacterial environment, but no precise measurement has yet been made. Fourth, the codons in ovalbumin mRNA are not randomly distributed[26] and their frequency may differ significantly from the frequency of codons in *E. coli*. Some codons (AUC, GUG, CCA, GAC, GAA, GCA) are used more frequently in ovalbumin than in coliphage genes of known sequence[36,37], such as MS2 or ΦX174. The possible deficiency of *E. coli* in the corresponding minor tRNAs does not seem to

interfere much with expression of the ovalbumin gene sequence.

This is, to our knowledge, the first example of synthesis by *E. coli* cells, of a long polypeptide chain from a higher eukaryote. The accuracy of the genetic construction which we made is not perfect, as five amino acids of the natural protein are missing and replaced by a few others from a bacterial protein. There is no theoretical reason, in our opinion, that the site of the junction between the *lac* and the ovalbumin sequences could not be moved down to the ATG initiator codon, leading to the synthesis of an accurate product. This is also the first example of a long polypeptide chain synthesised in good yield from a cloned complementary DNA made from mRNA. As a whole, our experiments demonstrate the feasibility of engineering bacteria for a potentially very large number of uses. Because of its well balanced amino acid content, ovalbumin is often used as a standard for nutritional value, and it is therefore conceivable that the ovalbumin-like product made by bacteria could be used for feeding purposes.

Biohazards associated with the experiments described in this publication have been examined previously by the French National Control Committee. The experiments were carried out accordingly in L1B1 conditions[38].

We thank Drs P. Charnay and P. Tiollais for supplying the DNA of phage λcI857p*lac*5ΔZL8UV5, Dr J. Rolland for help in the radioimmunoassays including the gift of rabbit antiserum, Mrs A. Hemard for her help in the preparation of the manuscript, Dr J. L. Mandel for providing the *Hha*ov fragment and the *Hha*ov ^{32}P-labelled probe, Drs J. P. Le Pennec, M. Cochet, and M. Lemeur, P. Farabaugh, M. Hofnung, A. Kornberg and F. Rougeon for the supply of materials, Dr R. Fauve for the use of his gamma-counter, Drs G. Brownlee and A. Fritsch for communication of unpublished results, and Drs C. Leguern and P. Chambon for discussion and information. This work was supported by grants from CNRS (ERA 201, plus an exceptional grant and ATP "Structure du génome") and from INSERM (CRAT 76.4.310 and 76.4.465). A.R. is a post-doctoral fellow of the National Cancer Institute of Canada.

Received 6 July; accepted 4 September 1978.

1. Struhl, K., Cameron, J. R. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1471–1475 (1977).
2. Struhl, K. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5255–5259 (1977).
3. Ratzkin, B. & Carbon, J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 487–491 (1977).
4. Clarke, L. & Carbon, J. *J. molec. Biol.* **120**, 517–532 (1978).
5. Vapnek, D., Hautala, J., Jacobson, J., Giles, N. & Kushner, S. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3058–3062 (1977).
6. Rambach, A. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5041–5045 (1977).
7. Polisky, B., Bishop, R. J. & Gelfand, D. M. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3900–3904 (1976).
8. Kourilsky, P., Gros, D., Rougeon, F. & Mach, B. *Nature* **267**, 637–639 (1977).
9. Itakura, K. *et al. Science* **198**, 1056–1063 (1977).
10. Jeffreys, A. J. & Flavell, R. A. *Cell* **12**, 1097–1108 (1977).
11. Tilghman, S. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 725–729 (1978).
12. Breathnach, R., Mandel, J. L. & Chambon, P. *Nature* **270**, 314–319 (1977).
13. Brack, C. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5652–5656 (1977).
14. Goodman, H. M., Olson, M. V. & Hall, B. D. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5453–5457 (1977).
15. Valenzuela, P. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 190–194 (1978).
16. Garapin, A. C. *et al. Nature* **273**, 349–358 (1978).
17. Garapin, A. C. *et al. Cell* (in the press).
18. Mandel, J. L. *et al. Cell* (in the press).
19. Doel, M. T., Houghton, M., Cook, E. A. & Carey, N. H. *Nucleic Acids Res.* **4**, 3701–3713 (1977).
20. Weinstock, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 1299–1303 (1978).
21. Fothergill, L. A. & Fothergill, J. E. *Biochem. J.* **116**, 555–561 (1970).
22. Palmiter, R. O., Gagnon, J. & Walsh, A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 94–98 (1978).
23. Lee, U. C. & Montgomery, R. *Archs Biochem. Biophys.* **97**, 9–17 (1962).
24. Milstein, C. P. *Biochem. J.* **110**, 127–134 (1968).
25. Wiseman, R. L., Fothergill, J. E. & Fothergill, L. A. *Biochem. J.* **127**, 775–780 (1972).
26. McReynolds, L. A. *et al. Nature* **273**, 723 (1978).
27. McReynolds, L. A., Catterall, J. F. & O'Malley, B. W. *Gene* **2**, 217–231 (1977).
28. Humphries, P. *et al. Nucleic Acids Res.* **4**, 2389–2407 (1977).
29. Rougeon, F., Kourilsky, P. & Mach, B. *Nucleic Acids Res.* **2**, 2365–2378 (1975).
30. Rougeon, F. & Mach, B. *J. biol. Chem.* **252**, 2209–2217 (1977).
31. Breathnach, R., Benoist, C., O'Hara, K., Gannon, F. X. & Chambon, P., in preparation.
32. Backman, K., Ptashne, M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **73**, 4174–4178 (1976).
33. Charnay, P., Louise, A., Perrin, D. & Tiollais, P. *Molec. gen. Genet.* (submitted).
34. Bolivar, F. *et al. Gene* **1**, 65–79 (1977).
35. Cami, B. & Kourilsky, P. *Nucleic Acids Res.* (in the press).
36. Fiers, W. *et al. Nature* **260**, 500–507 (1976).
37. Sanger, F. *Nature* **265**, 687–695 (1977).
38. *Prog. scient.* **191**, 91–94 (1977).
39. Dickson, R. C., Abelson, J., Barnes, W. M. & Reznikoff, W. S. *Science* **187**, 27–35 (1975).
40. Sharp, P. *et al. Biochemistry* **12**, 3055–3063 (1977).
41. Brown, N. L. & Smith, M. *FEBS Lett.* **65**, 284–287 (1976).
42. Smith, D. L., Blattner, F. R. & Davies, J. *Nucleic Acids Res.* **3**, 343–353 (1976).
43. Perricaudet, M. & Tiollais, P. *FEBS Lett.* **56**, 7–11 (1975).
44. Haggerty, D. M. & Schleif, R. E. *J. Virol.* **18**, 659–663 (1976).
45. Bolton, A. E. & Hunter, W. M. *Biochem. J.* **133**, 529–539 (1973).
46. Stephens, J. C., Artz, S. W. & Ames, B. N. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4389–4393 (1975).
47. Rhoads, R. E., McKnight, G. S. & Schimke, R. T. *J. biol. Chem.* **248**, 2031–2039 (1973).
48. Marsden, H. S., Crombie, I. K. & Subak-Sharpe, J. H. *J. gen. Virol* **31**, 347–372 (1976).
49. Studier, F. W. *J. molec. Biol.* **79**, 237–248 (1973).
50. Laskey, R. A. & Mills, A. D. *Eur. J. Biochem.* **56**, 335–341 (1975).
51. Ullrich, A. *et al. Science* **196**, 1313–1320 (1977).

Nucleotide sequence homology at 12 intron–exon junctions in the chick ovalbumin gene

J. F. Catterall & B. W. O'Malley
Dept of Cell Biology, Baylor College of Medicine, Houston, Texas 76706

M. A. Robertson, R. Staden, Y. Tanaka & G. G. Brownlee
MRC Laboratory of Molecular Biology, Hills Road, Cambridge, UK

A short partial sequence homology is present at all intron–exon junctions, or splice points, in the chick ovalbumin gene; it is probably a signal for a splicing enzyme. The significance of the junction sequences for splicing is discussed. We find no evidence of strong Watson–Crick base pairing between adjacent junctions.

OUR knowledge of the organisation of the ovalbumin gene has advanced rapidly because of the successful cloning of a major part of this gene[1–3]. Analysis of the clones[1] has shown that the gene is arranged in at least eight separate regions (I–VIII, Fig. 1), which are colinear with the mRNA sequence[4]. But these expressed pieces or exons[5] are separated by seven sections of DNA or introns labelled A–G which are absent in the mature mRNA. Given this remarkably complex organisation, how is the mRNA synthesised. Potentially an RNA polymerase could transcribe only the exons leaving the introns alone. This 'jumping polymerase' model[6] would essentially join up sections I–VIII to give the correct mature mRNA. Alternatively the RNA polymerase could synthesise a copy of the entire gene copying both intron and exon regions. This pre-mRNA could be cut and the exons joined in the correct order, the introns being eliminated. The latter hypothesis seems more likely as there is evidence of a 40S pre-mRNA[7] containing both exon and intron sequences. This mechanism also occurs for other spliced genes such as those present in adenovirus[8–10], the β-globin[11] and yeast tRNA genes[12,13].

The splicing must be a precise process. It has been suggested that two adjacent exons distant in an immunoglobulin λ chain

TAB M

occur (Fig. 7*b*). Subsequently inverted repeat pair 5 of the newly synthesised DNA strand could fold back on itself forming a new template for further DNA synthesis (Fig. 7*c*). The polymerase could then proceed in the 3′ direction along the newly synthesised strand until it passes inverted repeat pair 3 and reaches position 33 (Fig. 7*d*). At this point the same slippage process is repeated: melting out, folding back of the inverted repeat pair 3 and DNA synthesis along the newly made DNA strand leading to the structure shown in Fig. 7*e*. This slippage tends to form long inverted repeats. Occasionally these may resist partial melting out so that DNA synthesis can then proceed along the old parental DNA strand, provided that inverted repeat pair 3 (33–52) in the parental DNA strand has folded up as indicated in Fig. 7*e* to yield a stable structure.

On another round of replication the DNA heteroduplex molecule forms two segregants, the old parental IS2 sequence and the newly generated IS2–6 allele.

The slippage of the DNA template during replication of IS2 as outlined in Fig. 7 generates symmetrical DNA additions (see Fig. 5). The process requires, first, a DNA region rich in A–T, and second, pairs of inverted repeat sequences at both ends of the A–T rich region. The newly formed sequences generated by this process may contain new signals for the turn-on of gene expression, as shown by IS2–6. Which parts of the IS2–6 sequences constitute the turn-on signal is not yet known.

The replication 'slippage' model described in Fig. 7 for IS2–6 allows us to predict the DNA sequence of the replication 'slippage' product of the upper DNA strand of IS2. Since preparation of this manuscript we have determined the sequence of the 54-base pair mini-insertion IS2–7. Its sequence is identical to the slippage product predicted for replication of the upper strand of IS2 and will be reported elsewhere[13].

We thank Gudrun Hoeksma for technical assistance and Patricia Nevers for reading the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft.

Received 8 May; accepted 4 August, 1978.

1. Saedler, H. & Heiss, B. *Molec. gen. Genet.* **22**, 267–277, (1973).
2. Starlinger, P. & Saedler, H. in *Curr. Topics Microbiol. Immun.* **75**, 111 (1975).
3. Saedler, H., Reif, H. J., Hu, S. & Davidson, N. *Molec. gen. Genet.* **132**, 265–289 (1974).
4. Hlacinsky, W. *et al. Gene* **2**, 61–74 (1977).
5. Ghosal, D. & Saedler, H. *Molec. gen. Genet.* **158**, 123–128 (1977).
6. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
7. Musso, R. E. & Rosenberg, M. in *DNA Insertion Elements, Plasmids, and Episomes* (eds Bukhari, A. I., Shapiro, J. A. & Adhya, S. L.) 597–600 (Cold Spring Harbor, 1977).
8. Streeck, R. E. & Hobom, O. *Eur. J. Biochem.* **57**, 595–606 (1975).
9. Thuring, R. W. J., Sanders, J. P. M. & Borst, P. *Analyt. Biochem.* **66**, 213–221 (1975).
10. Brack, Ch., Eberle, H., Bickle, T. A. & Yuan, R. *J. molec. Biol.* **10**, 305–309 (1976).
11. Musso, R., Di Lauro, R., Rosenberg, M. & de Crombrugghe, B. *Proc. natn. Acad. Sci. U.S.A.* **74**, 106–110 (1977).
12. Schlesinger, O. H., Niall, H. D. & Wilson, D. *Biochem. biophys. Res Comm.* **61**, 282–289 (1974).
13. Ghosal, D. & Saedler, H. *Cold Spring Harb. Symp. quant. Biol.* **43**, (in the press).

Phenotypic expression in *E. coli* of a DNA sequence coding for mouse dihydrofolate reductase

Annie C. Y. Chang*†§, Jack H. Nunberg‡, Randal J. Kaufman‖, Henry A. Erlich†, Robert T. Schimke‡ & Stanley N. Cohen*†¶

Departments of Genetics*, Medicine†, Biological Sciences‡, Medical Microbiology§, and Pharmacology‖, Stanford University, Stanford, California 94305

The construction and analysis of bacterial plasmids that contain and phenotypically express a mammalian genetic sequence are described. Such plasmids specify a protein that has enzymatic properties, immunological reactivity and molecular size characteristic of the mouse dihydrofolate reductase, and render host cells resistant to the antimetabolic drug trimethoprim.

SINCE the initial propagation of eukaryotic DNA in bacteria[1], several systems have been used to study the expression in *Escherichia coli* of DNA derived from higher organisms. Biological activity of genes from the lower eukaryotes, *Saccharomyces cerevisiae*[2,3] and *Neurospora crassa*[4], has been demonstrated using phenotypic selection for functions that complement mutationally inactivated homologous bacterial genes. Immunological reactivity with antibody made against human somatostatin was shown for a peptide fragment cleaved *in vitro* from a hybrid protein encoded in part by bacterial DNA and in part by a chemically synthesised somatostatin DNA sequence[5]. Very recently, a protein containing amino acids of rat proinsulin was shown to be made by bacteria that carry a double-stranded complementary DNA (cDNA) transcript of pre-proinsulin mRNA[6]; in that instance, antigenic determinants for both insulin and the bacterial enzyme β-lactamase were detected on a fused peptide transported outside the cell. It is not known, however, whether the mammalian peptide components of such immunologically reactive hybrid proteins have functional biological activity.

Our approach to the study of mammalian gene expression in bacteria has been to generate a heterogeneous population of clones carrying a DNA sequence that codes for a selectable mammalian gene product, and then to select directly those bacteria in the population that phenotypically express the genetic sequence. The mammalian enzyme dihydrofolate reductase (DHFR), which catalyses the conversion of dihydrofolic acid to tetrahydrofolic acid, is especially suitable for this purpose. The mammalian DHFR has a much lower affinity for the antimetabolic drug, trimethoprim (Tp), than does the corresponding bacterial enzyme[7]. Thus, bacteria which biologically express mammalian DHFR activity are resistant to levels of trimethoprim that ordinarily inhibit growth.

When these studies were initiated, the only bacterial host approved for EK2 recombinant DNA experiments[8] was *E. coli* K12 strain χ1776 (ref. 9). As this strain is already resistant to

¶To whom correspondence should be addressed at Room S337, Stanford University Medical Center, Stanford, California 94305.

high concentrations of trimethoprim because of its *thy*⁻ mutation[10], direct selection of bacteria that synthesise the mammalian DHFR could not be carried out. Therefore, our initial studies used an *in situ* hybridisation procedure[11] to identify χ1776 clones which carried a mouse DHFR cDNA sequence, with the expectation that a highly sensitive indirect radioimmunoassay[12] would be used to detect any clones that expressed antigenic determinants of the protein product encoded by the gene.

Construction and cloning of chimaeric plasmids containing a DHFR ds cDNA

Figure 1 summarises the experimental scheme used in this study. Partially purified mRNA containing DHFR sequences from methotrexate-resistant AT-3000 mouse cells[13] was used in the preparation of double-strand (ds) cDNA by RNA-dependent DNA polymerase (reverse transcriptase) and DNA polymerase 1 (ref. 14 and Fig. 2). Homopolymeric deoxy-C 'tails' were added to the unfractionated cDNA by terminal deoxynucleotidyl transferase (Fig. 2) and the section of the gel containing the predominant (1,500 base pairs) tailed cDNA was eluted. The material recovered from the gel was annealed with an equimolar concentration of pBR322 plasmid DNA that had been cleaved in the β-lactamase gene by the *Pst*I endonuclease[15] and treated with terminal transferase to add homopolymeric dG tails at the cleavage sites. The extent of annealing of DHFR ds cDNA with the plasmid vector was monitored by electron microscopy[16] using circle formation as an indicator. *Pst*I sites are regenerated at both ends of the insert as a result of such re-circularisation[17,18].

Constructed chimaeric plasmids were introduced into $MnCl_2$-treated[19] *E. coli* K12 strain χ1776 and tetracycline (Tc)-resistant transformants (yield ~30 colonies per ng DNA) were selected and tested separately for the presence of a DNA species complementary to a highly purified DHFR cDNA probe. About 40% of the Tc-resistant clones gave a positive reaction by an *in situ* hybridisation test[11] (Table 1 and Fig. 3). Plasmids from 14 of the reacting clones were examined by gel electrophoresis and all were found to contain a single DNA insert approximately 1,500 base pairs in length. As both ends of the *Pst*I-cleaved pBR322 DNA receive homopolymeric dG tails, only those molecules that have acquired a dE-tailed cDNA insert or had escaped cleavage and/or dG tailing would be expected to re-circularise and transform; thus, we presume that most of the nonreacting

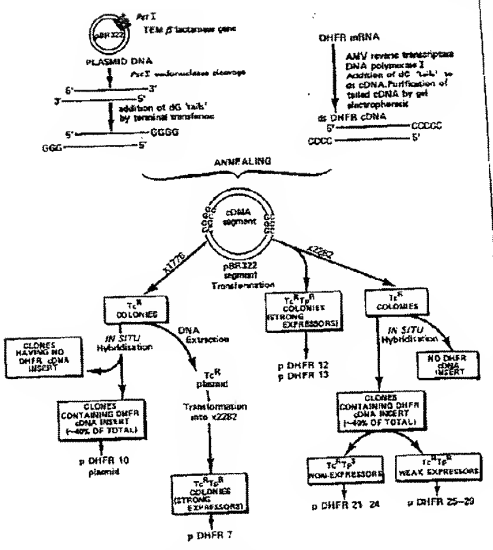


Fig. 1 Scheme used for cloning and expression of mouse DNA sequences that code for DHFR. Additional experimental details are given in the legend to Table 1 and in the text.

Tc-resistant colonies contain inserts of contaminating non-DHFR cDNA.

When χ2282, a *thy*⁺ variant of χ1776, was approved for EK2 use, direct selection of bacteria that phenotypically expressed the eukaryotic DNA sequence became feasible (see Fig. 1). Two populations (25 colonies each) of previously identified Tc-resistant colonies of χ1776 were pooled, and plasmid DNA extracted from these populations was introduced by transformation into χ2282. In other experiments, we transformed χ2282 directly with annealed pBR322–DHFR cDNA (Table 1). Colonies that expressed resistance to both Tc and Tp were obtained in both types of experiments; three independent

Table 1 Transformation experiments using χ1776 and χ2282

Host	Transforming DNA: Plasmid vector	Transforming DNA: Insert	Transformants per ng DNA: Tc (5 μg ml⁻¹)	Tc (10 μg ml⁻¹)	Tc/Tp (5 μg ml⁻¹ of each)	*In situ* hybridisation with DHFR cDNA probe (% positive)
χ1776	pBR322	None	4×10^{2}	2×10^{2}	$<2\times10^{-3}$	—
χ1776	pBR322	cDNA(1°)	60	32	—	40
χ2282	pBR322	cDNA(1°)	70	—	2	44
χ2282	pBR322	cDNA(2°)	60	—	1.3×10^{-1}	—
χ2282	pDHFR7		75	—	25	—

pBR322 plasmid DNA that had been annealed *in vitro* with dC-tailed DHFR cDNA (designated 1°) was introduced into χ1776 or χ2282, using a modification of a previously described transfection procedure[19]. 1 ml of an overnight bacterial culture was inoculated into 100 ml of L broth supplemented with diaminopimelic acid (DAP, 50 μg ml⁻¹) and (for χ1776 only) thymidine (4 μg ml⁻¹). Bacterial cultures were grown until exponential phase at 35°C and then collected by centrifugation at 4°C. Cells were washed in 0.3 volume 10 mM NaCl, resuspended in 30 ml freshly prepared MCN buffer (70 mM $MgCl_2$; 40 mM sodium acetate, *p*H 5.6, and 30 mM calcium chloride) and chilled on ice for 20 min. Cells were collected, resuspended in 1 ml MCN and added in 200-μl aliquots to 50 μl DNA in TEN (10 mM Tris-HCl, *p*H 7.5, 0.1 mM EDTA, 50 mM NaCl) or MCN buffer. After chilling at 0° C for 30 min, reactions were incubated at 27 °C for 5 min, chilled again for 30 min, and 50-μl samples were plated on to Penassay broth agar supplemented with DAP, thymidine (for χ1776), and antibiotics as indicated. When χ2282 was used, the selective medium was M9 minimal agar supplemented with 0.5% casamino acids, biotin (2 μg ml⁻¹), DAP (50 μg ml⁻¹) and Tp (2.5–10 μg ml⁻¹) plus tetracycline (Yc) or kanamycin (Km) as indicated. Plates containing transformants were incubated at 32 °C and colonies were scored 2–3 d after plating. pBR322 plasmid DNA lacking the cDNA insert was used as a control. cDNA preparations labelled as 2° consisted of plasmid DNA isolated from a nonfractionated population of clones that had previously been transformed with chimaeric molecules carrying a cDNA insert. In the experiment shown in the last line the transforming DNA was isolated from a clone (pDHFR7) that expresses resistance to Tp as well as Tc.

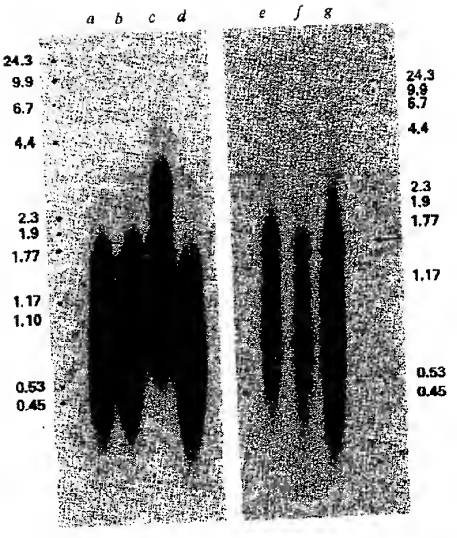


Fig. 2 Preparation and characterisation of DHFR cDNA insert DNA complementary to DHFR mRNA was synthesised essentially as described elsewhere[14], using avian myeloblastosis virus (AMV) reverse transcriptase and polysomal RNA obtained by indirect immunoprecipitation of DHFR-synthesising polysomes from methotrexate-resistant AT-3000 S-180 mouse cells[13]. The RNA had been estimated to contain DHFR mRNA as 20% of its mRNA[13]. The reaction was carried out in 100 μl using 340 μg polysomal RNA (estimated to contain 5 μg polyA–RNA), 45 units AMV reverse transcriptase and dCTP labelled to 4 Ci mmol^{-1} with ^{32}P-dCTP (Amersham). Approximately 1.4 μg cDNA was synthesised in 30 min. The reaction was stopped by the addition of EDTA (to 10 mM) and extracted with phenol, followed by ether, before being passed over a Sephadex G-50 fine column in 10 mM Tris, *p*H 7.4, 2 mM EDTA and, 10 mM NaCl (TEN). The void volume was collected and precipitated with ethanol. After centrifugation, the RNA was hydrolysed with NaOH (ref. 31), neutralised and precipitated with ethanol. The cDNA was then used as template for the synthesis of the second strand by *E. coli* DNA polymerase I essentially as described[31]. The reaction took place at 42 °C for 10 min in 100 μl using 1.1 μg cDNA, 10 units DNA polymerase I, and 200 μM of each deoxynucleotide triphosphate with dCTP adjusted to 30 Ci mmol^{-1} as above. Approximately 0.85 μg of the second strand was synthesised. The reaction was stopped and extracted as above before being passed over a Sephadex G-50 fine column in TEN containing only 0.1 mM EDTA. Column fractions containing the ds cDNA were then treated with *Aspergillus oryzae* S_1 nuclease as described elsewhere[31]. After extraction and precipitation with ethanol, approximately 1.0 μg ds cDNA was obtained. Aliquots of *a*, first strand product; *b*, first strand product after base treatment; *c*, second strand product and *d*, second strand product after S_1 nuclease treatment were examined on a 1.5% agarose gel in alkaline conditions as described[32]. Terminal addition of dCTP to the ds cDNA by terminal deoxynucleotidyl transferase (TdT), prepared as described elsewhere[33] was carried out by a modification[34] of the Co^{2+} procedure[35]. The reaction was carried out in 500 μl containing 140 mM cacodylic acid, 30 mM Tris base, 110 mM KOH (final *p*H 7.6), 0.1 mM dithiothreitol, 150 μM dCTP (adjusted to 8 Ci mmol^{-1} with ^{3}H-dCTP (Amersham)), 1 mM $CoCl_2$ (added to prewarmed reaction mix before enzyme addition), approximately 1.0 μg ds cDNA (assuming an average MW giving approximately 600 base pairs, this provides 10 pM 3′ termini per ml) and 0.5 μl TdT (2.3×10^5 units ml^{-1}). The reaction was allowed to proceed at 37 °C for 10 min before being cooled and sampled to determine incorporation. Approximately 30 dC residues were added per 3′ terminus. The reaction was stopped, extracted, desalted and precipitated with ethanol as above. Aliquots of *e*, second strand product; *f*, second strand product after S_1 nuclease treatment, and *g*, dC-tailed ds cDNA were analysed on a 1.7% agarose gel in Tris-acetate-NaCl (ref. 36). The dC-tailed ds cDNA was then preparatively electrophoresed on a similar gel and the '1,500-base pair' region (arrow) cut out of the gel and electrophoretically eluted into a dialysis bag as described elsewhere[37]. The eluted material was extracted as above, concentrated by lyophilisation and precipitated with ethanol. After centrifugation, the 1,500-base pair dC-tailed ds cDNA (approximately 80 ng) was redissolved in 10 mM Tris HCl, *p*H 7.4, 0.25 mM EDTA and 100 mM NaCl (annealing buffer). pBR322 plasmid DNA, isolated as described elsewhere[38] was digested with a 1.5-fold excess of *Pst*I endonuclease in conditions suggested by the vendor (New England Biolabs) and the linear plasmid DNA was cut out and eluted from a 0.7% agarose gel in TBE[39] as described above. The plasmid DNA was 'tailed' with dG residues using procedures similar to those described above. Approximately 15–20 dG residues were added per 3′ terminus. Following extraction, the dG-tailed vector was passed over a Sephadex G-50 fine column in annealing buffer and the void volume was collected. Equimolar amounts of dC-tailed ds cDNA and dG-tailed vector DNA were allowed to anneal essentially as described by W. Rowe and R. A. Fratel (personal communication) except that the vector concentration was kept at 75 ng ml^{-1} in the annealing reaction. Circularisation was monitored by electron microscopy[16] and was typically about 20–40%. This annealed DNA was used directly for transformation into χ1776 or χ2282.

transformants were found to be capable of immediate growth in media containing at least 1,000 μg ml^{-1} of Tp, and were termed 'strong expressors'. Plasmid DNA isolated from these colonies was shown by repeat transformation to encode both the Tc and Tp resistance phenotypes. However, Tp^RTc^R transformants occurred in different experiments at only 20–60% of the frequency observed when selection of transformants was carried out for Tc resistance alone. Transfer of Tc-selected colonies on to minimal medium agar plates containing Tp showed that all such clones express resistance to at least 1,000 μg ml^{-1}. Together, these findings suggest that phenotypic expression of Tp resistance may be delayed in transformants until a sufficient quantity of plasmid-specified DHFR has accumulated. Analogous results have been obtained with other antimicrobial drug-resistance determinants encoded by plasmids introduced into *E. coli* by transformation[20].

Structure of the DHFR cDNA

Gel electrophoresis of endonuclease-cleaved plasmid DNA from three separately derived bacterial clones that expressed high levels of trimethoprim resistance showed similar overall patterns. Using such data, a cleavage map (Fig. 4) of the cDNA insert of one of these (pDHFR7) was constructed.

Examination of the amino acid sequence of the mouse DHFR enzyme[21] allowed us to assign one of the *Hae*III cleavage sites to the *trp-pro* (TGG–CCX) present at amino acid positions 24 and 25, thus localising the DHFR structural sequence to a position near the 5′ end of the mRNA template used in synthesising the cloned cDNA. Other cleavage sites within the DHFR structural sequence were consistent with positions predicted by computer analysis of the amino acid sequence. DNA sequence analysis of two separate regions of the cDNA provided direct verification that the nucleotide sequence of the insert corresponds to the amino acid sequence reported for the mouse DHFR enzyme, and also confirmed that the coding sequence for DHFR is located at the ds cDNA equivalent of the 5′ end of the mRNA.

The nucleotide sequence at the pBR322–cDNA junction nearest the 5′ end of the mRNA used as template for DHFR cDNA (that is, at the *Pst*I site) is of special interest. The complement of the 'sense' strand of β-lactamase gene of the vector (J. G. Sutcliffe, personal communication) is interrupted at the *Pst*I site by a series of 11 dG residues added by the terminal transferase, and these are followed immediately by (1) an ATG (AUG) protein start codon, and (2) the codon for the first amino acid of the mouse DHFR structural gene. The sequence that codes for the mouse DHFR is in the same orientation as at that encoding the β-lactamase on the vector plasmid; however, the

number of incremental G residues (that is, 10) at the vector–insert junction ensures that the DHFR cDNA sequence is not in the same translational reading frame as the β-lactamase gene.

If the phenotypic expression we have observed for the mouse DHFR sequences in bacteria is the result of translational readthrough from signals that initiate protein chains within the β-lactamase gene, then the host bacterial cell must be able to circumvent the observed frame shift by a 'slippage' mechanism of translation. Such an event might potentially be aided by the long run of dG residues introduced at the pBR322–cDNA junction and could account for our observation (Table 1) that plasmids containing a DHFR cDNA insert yield more than twice as many expressors as would be expected from considerations of reading frame and orientation. However, the high level of functional expression observed for both primary and secondarily transformed clones of pDHFR7 does not seem to be readily explained by slippage of tRNA molecules during translation.

Perhaps a more likely explanation is that the *Pst*I–polyG–ATG sequence that has been constructed preceding the coding sequence for DHFR serves as a binding and protein initiation site for the bacterial ribosome. Recent studies[22,23] have identified sequences on mRNA in the 5′ direction from the initiator codon that are complementary to the CCUCC sequence at the 3′ end of the 16S ribosomal RNA species, proposed by Shine and Dalgarno[24] to be involved in the binding of mRNA to ribosomes. It is tempting to speculate that the mRNA transcript from the sequence at the pBR322–cDNA junction has sufficient complementarity to the CCUCC sequence to allow ribosomal binding when a translational start signal is located an appropriate distance away. In such an event, the ATG protein start signal that immediately precedes the coding sequence for the mouse DHFR might initiate a peptide chain having a size characteristic of the mammalian enzyme. Immunological analysis of extracts derived from pDHFR7 and other expressing clones has yielded results consistent with this interpretation (see below).

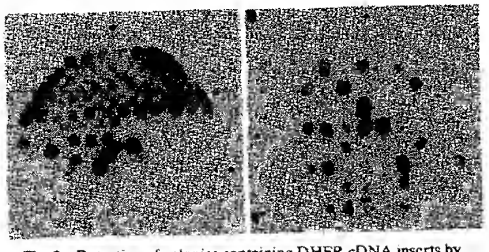


Fig. 3 Detection of colonies containing DHFR cDNA inserts by *in situ* hybridisation. Colonies were screened for DHFR sequences using a modification (G. N. Dnell, unpublished) of an *in situ* hybridisation procedure[11]. Tc-resistant colonies were transferred to nitrocellulose filters (Millipore, HAWG) that had been placed on Penassay broth agar plates containing Tc (10 μg per ml). (Filters had been washed twice by boiling in H_2O and autoclaved before being placed on plates.) After 2–3 d of bacterial growth at 32°C, the filter was removed from the plate and placed on a Whatman no. 3 pad saturated with 0.5 M NaOH. After 7 min, the filter was sequentially transferred to a series of similar pads saturated with 1 M Tris, *p*H 7.4 (twice, 7 min each); 1.5 M NaCl, 0.5 M Tris, *p*H 7.5 (once, 7 min); and 0.30 M NaCl, 0.03 M Na citrate (2×SSC) (once, 7 min) After the excess liquid had been removed by suction, the filter was placed on a pad containing 90% ethanol, dried by suction and baked *in vacuo* at 80°C for 2 h. Before hybridisation, filters were pretreated for 3–6 h at 65°C in hybridisation buffer that contained 5×SSC, *p*H 6.1 0.2% SDS, 0.02% Ficoll 400 (Pharmacia) and 8 μg ml^{-1} *E. coli* tRNA. Hybridisations were carried out with individual filters in 1.5 ml hybridisation buffer containing 2×10^4 c.p.m. ^{32}P-labelled purified DHFR cDNA[15] in a sealed plastic bag at 65 °C for 24 h. The filters were then washed in hybridisation buffer (once, 60 min at 65 °C); in 5×SSC, *p*H 6.1 (three times, 60 min each at 65 °C); and in 2×SSC, *p*H 7.4 (twice, 10 min at room temperature), air dried, and prepared for autoradiography. Left: top, a collection of χ1776 colonies which contain a DHFR cDNA insert; middle, Tc-resistant χ2282 colonies derived from transformation with annealed pBR322 ds cDNA—both reacting and non-reacting colonies are seen; bottom, colonies containing pBR322 and pACYC101 plasmids which show no visible hybridisation. Right: several positive colonies on a representative filter analysed in screening χ1176 transformants. Negative colonies represent clones containing pBR322 or pACYC101 plasmids which show no visible hybridisation.

Analysis of enzyme activity encoded by pDHFR7 plasmid

Mammalian dihydrofolate reductases can be distinguished from their bacterial counterparts by the ability of mammalian enzymes to use folate as a substrate and by their differential sensitivity to competitive inhibitors[25,26]. In initial experiments, the reduction of folate to tetrahydrofolate was measured using extracts from the pDHFR7 clone, from a Tp-sensitive clone containing a DHFR cDNA insert (the pDHFR10 plasmid), and from cells that contain only the pBR322 vector. Although all three clones are capable of synthesising a chromosomally produced bacterial enzyme, only the enzyme present in extracts from cells containing the pDHFR7 plasmid gave reduction of folate (4×background). Additional evidence that the reductase encoded by the pDHFR7 plasmid is of mammalian origin was obtained by inhibitor analysis (Fig. 5). The DHFR isolated directly from mouse cells and the activity encoded by the pDHFR7 plasmid showed identical sensitivities to methotrexate, trimethoprim and a triazine derivative (2,4,-diamino-1-(4′-butylphenol)-6,6 dimethyl-1, 6-dihydro-1,3,5-triazine); both enzyme activities were 200 times more sensitive to the triazine than to trimethoprim. In contrast, bacterial dihydrofolate reductase is inhibited more effectively by trimethoprim ($K_i = 5\times10^{-9}$) than it is by triazine ($K_i = 6.5\times10^{-4}$)[25].

As methotrexate binds stoichiometrically to dihydrofolate reductase[27], we can estimate the number of molecules of enzyme in the pDHFR7 plasmid extracts from the methotrexate inhibition date of Fig. 5. We calculate from the specific activity of the extract (3 units per mg of soluble protein) and the specific activity and methotrexate binding parameters of the mouse enzyme[28] that 0.01% of the soluble bacterial protein is active mammalian DHFR.

Immunological characterisation of bacterial cell extracts containing mouse DHFR

Immunological evidence confirming the nature of the DHFR encoded by pDHFR7 and other plasmids that contain a mouse DHFR cDNA insert was obtained using a solid-phase sandwich radioimmunoassay[12]. Tp^R clones of χ2282 containing the independently derived plasmids pDHFR 7, 12 and 13 showed a strong reaction with rabbit antibody directed against mouse DHFR in an *in situ* immunoassay[12] (data not shown); protein that reacted with the antibody was also made by bacteria which showed low levels of phenotypic expression and by some clones that did not make a biologically functional DHFR (that is, were Tp sensitive). The nature of the antigen synthesised by Tp^S and Tp^R clonnes was examined more fully using a newly developed method (filter affinity transfer, or FAT procedure) for the *in situ* immunological characterisation of proteins in gels[29]. This procedure depends on the covalent coupling of F(ab)$_2$ antibody fragments to a chemically derivatised and activated cellulose filter; antigen transferred on to the filter from an SDS-polyacrylamide gel is detected by subsequent incubations with antiserum and ^{125}I-labelled *Staphylococcus aureus* protein A (ref. 12).

Filter affinity transfer analysis of the pDHFR7 extract (Fig. 6, lanes *b*, *c*) shows the presence of protein that reacts immunologically with the antibody to mouse DHFR and further shows

CCUCC
onal start
an event,
:edes the
a peptide
enzyme.
[FR7 and
with this

hed from
mmalian
fferential
:riments,
·ed using
ve clone
nid), and
ıough all
osomally
ı extracts
action of
eductase
·igin was
isolated
l by the
:thotrex-
10-1-(4'-
e); both
: triazine
lrofolate
ım (K_i =

lrofolate
' enzyme
e inhibi-
ty of the
specific
mouse
is active

rial

: DHFR
a mouse
andwich
ing the
howed a
t mouse
protein
ia which
e clones
is, were
Tp^s and
veloped
ie *in situ*
his pro-
ıntibody
:ellulose
S–poly-
ns with
otein A

: (Fig. 6,
nmuno-
r shows

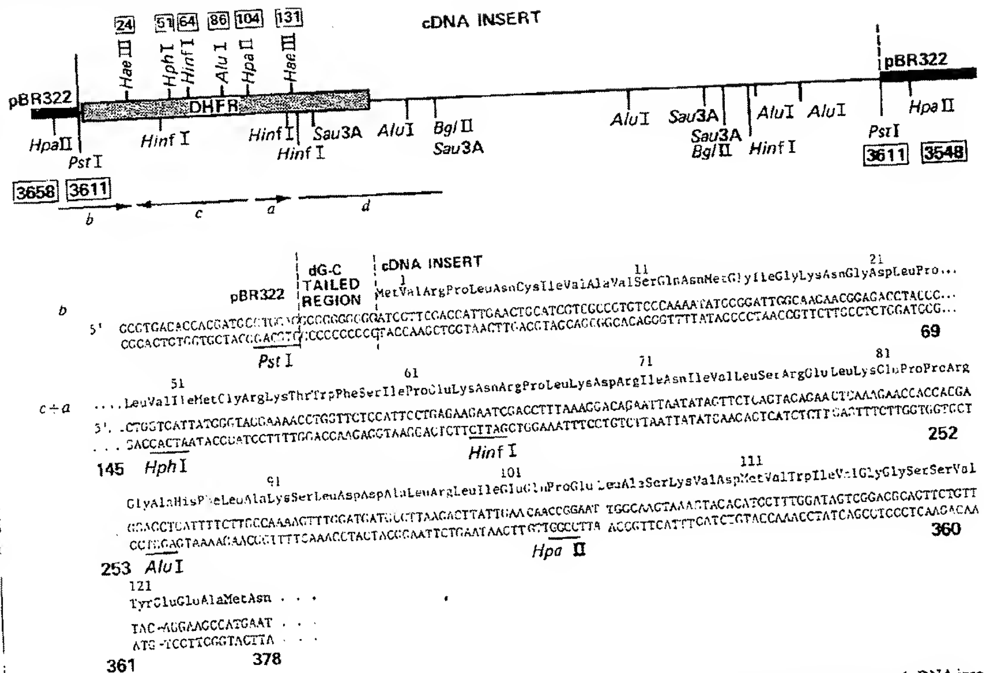


Fig. 4 Map of cDNA insert and adjacent regions of pDHFR 7 plasmid. Endonuclease cleavage map and partial DNA sequence of cDNA insert of the pDHFR 7 plasmid. The shaded area indicates the structural sequence for mouse DHFR. The locations of endonuclease cleavage sites were determined from polyacrylamide gel electrophoresis patterns following simultaneous or sequential digestions of either intact pDHFR or fragments isolated after earlier digestions. Restriction endonucleases (New England Biolabs or Bethesda Research Laboratories) were used according to the vendor's recommendations. Cleavage sites listed above the shaded area were assigned to specific amino acid positions within the mouse enzyme by nucleotide sequence. The boxed numbers of the *Hpa*II and *Pst*I sites indicate the locations of the sites in the pBR322 plasmid as determined by Sutcliffe (personal communication), and were used in orientating the amino acid sequence of the DHFR gene with respect to the β-lactamase sequence. The nucleotide sequence in the vicinity of the pBR322–cDNA junction corresponding to the 5' end of the mRNA used as template for the cDNA, and the sequence in the region of the *Hpa*II site at amino acid 104 of the structural sequence are shown. Nucleotides and amino acids are numbered from the start of the DHFR coding sequence; nucleotides in the 5' direction on the mRNA from position 1 have negative numbers. The DNA sequence shown were determined by the method of Maxam and Gilbert[40]. Fragments ($d+c$) and ($a+b$) were 5' end-labelled at the *Hpa*II and *Bgl*II-generated ends, treated with *Hae*III endonuclease, and subjected to electrophoresis in 6% acrylamide gel and TBE buffer[39]. Fragments *a*, *b* and *c* were eluted from the gel[32], precipitated with ethanol and used directly for base sequence determination using the gel system described in refs 40 and 41.

that most of the immunologically reactive material has the same electrophoretic mobility as enzyme obtained directly from mouse cells (molecular weight 22,000) (Fig. 6, lane *a*). An immunologically reactive band that migrates at this position was also seen in the material eluted with folic acid from a methotrexate affinity column that contained an extract from χ2282 (pDHFR7) cells (Fig. 6, lane *d*), suggesting that the 22,000 MW protein made by these bacterial cells has binding sites for both methotrexate and folate. Additional immunologically reactive bands which have mobilities consistent with a MW of 30,000–90,000 were seen in varying amounts in different extracts of bacterial cells that contained pDHFR7 (Fig. 6, lanes *b*, *c*); as the 1,500-base pair insert in this plasmid is capable of coding for a polypeptide no larger than 50,000 MW, we conclude that the most slowly moving bands are likely to be hybrid proteins that include antigenic sites of the mouse DHFR. Immunologically reactive high MW proteins are also made by a Tp^S clone (pDHFR21, Fig. 6, lane *e*) which contains a DHFR cDNA insert; however, this clone fails to synthesise an immunologically reactive 22,000 MW band. No immunological reactivity with antibody made to mouse DHFR was detected in extracts of cells carrying only the pBR322 vector (Fig. 6, lane *f*).

Variation in level of expression of mouse DHFR cDNA sequences in bacteria

Thirty-two separately derived clones of χ2282 transformants that had been selected on plates containing only Tc were replated on medium containing both Tc and Tp, and also were tested for the presence of a DHFR cDNA insert by *in situ* hybridisation. Fourteen colonies contained sequences homologous with the purified DHFR cDNA probe, and five of these (termed 'weak expressors') grew on plates containing 5 $\mu g\ ml^{-1}$ or more Tp. The remaining nine clones that contained DHFR sequences failed to show any growth on concentrations of Tp above 2.5 $\mu g\ ml^{-1}$, and were termed 'non-expressors'. Plasmid DNA isolated from four of the nine non-expressors and from all five weak expressors was analysed by gel electrophoresis, and the mean inhibitory concentration (MIC) of Tp was determined for each of the clones. The structural relationship of the cDNA

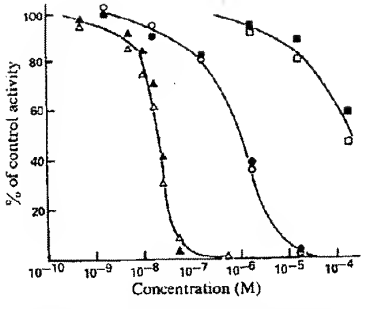


Fig. 5 Inhibitor analysis of DHFR from bacterial cells. Stationary phase cultures of χ2282 expressing trimethoprim resistance were grown in the presence of Tp (1 μg ml^{-1}) in minimal medium, washed with isotonic saline, and suspended in 50 mM potassium phosphate buffer, *p*H 7.0, containing 10 mM benzamidine and 10 mM phenyl methyl sulphonyl fluoride (3 volumes buffer to 1 volume cells). The suspension was sonicated and centrifuged at 10,000 r.p.m. for 15 min. The supernatant was centrifuged for 1 h at 100,000*g* before being studied. An R_2 methotrexate-resistant mouse cell extract was prepared as described elsewhere[42]. Enzyme activity was measured by the radioactive folic acid assay previously described[28]. Protein was determined by the method of Lowry[43]. Approximately 3 units of activity from the χ2282 extract or 5 units from the methotrexate-resistant mouse cell extract were incubated with inhibitor for 10 min at 24°C before assaying for folate reductase activity; the concentrations shown represent the final concentration of inhibitor in the reaction mixture. Background values, determined by measuring enzyme activity in the presence of 10 mM methotrexate, have been subtracted from all points. The results presented are the average of duplicate samples which generally varied by less than 10% and are expressed as a percentage of the value obtained in the absence of inhibitor. One unit of activity is the amount of enzyme needed to reduce 1 nmol of folate in 15 min at 37°C. △ and ▲ indicate addition of methotrexate, ○ and ● indicate addition of the triazine derivative, and □ and ■ indicate trimethoprim addition for χ2282 and the mouse cell extracts, respectively.

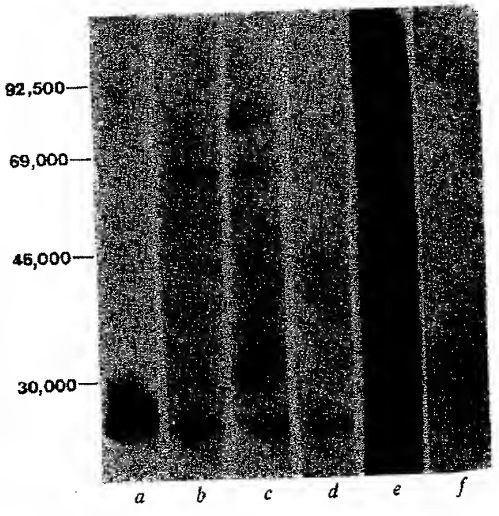
insert to the β-lactamase gene sequence of the vector, the length of each insert, and the minimal inhibitory concentration (MIC) determined for the clone are shown in Fig. 7. As can be seen, plasmids pDHFR7, 12, 13 and 26–29 all contain a complete DHFR structural sequence inserted in the same orientation (that is, orientation a) as the gene encoding the bacterial β-lactamase. The clone carrying each of these plasmids expresses Tp resistance, although the MIC varies from 150 μg ml^{-1} for pDHFR28 to >1,000 μg ml^{-1} for pDHFR7, 12 and 13, the greatest reactivity with antibody to mouse DHFR occurs with pDHFR12 (unpublished data).

It is unlikely that translational reading frame is the determining factor in the different levels of expression observed in these clones, as our DNA sequence analysis indicates that a correct reading frame is not essential for efficient expression in pDHFR 7, 12 or 13. However, the positioning of the putative ribosomal binding site in relation to the ATG start codon may potentially influence the strength of expression by affecting the formation of the translational initiation complex (compare with ref. 23) (Figs 4, 7).

The end of DHFR structural sequence that corresponds to the 5′ end of the mRNA is not present in plasmids pDHFR 21 and 24, thus explaining the observed lack of functional expression of the cDNA in these clones. However, pDHFR21, which has lost less than 15% of the structural sequence, nevertheless encodes a (probably hybrid) peptide that contains antigenic sites which react with antibody to mouse DHFR (Fig. 6 and unpublished data). It is particularly interesting that the clone carrying pDHFR25 expresses a low level of Tp resistance, although the coding sequence for the DHFR enzyme is inserted in an orientation opposite to that of the β-lactamase gene. This finding, and our detection in extracts of the pDHFR25 clone of protein that reacts immunologically with antibody to the mouse enzyme (unpublished data), suggest that readthrough transcription into the DHFR coding sequence from a promoter sequence located on or near the distal segment of the β-lactamase gene may occur. Consistent with this interpretation are preliminary data suggesting that DHFR antigenic sites are also synthesised by cells carrying pDHFR23, which is a non-expressor of Tp resistance and contains a cDNA insert in orientation b (Fig. 7). Further study is required to determine whether sequences in the distal segment of the β-lactamase gene are capable of serving as weak promoters for the initiation of mRNA chains that extend into the DHFR cDNA.

The findings reported here indicate that the bacterial clones we have constructed are synthesising and phenotypically expressing DHFR encoded by mouse cDNA sequences: (1) the cDNA insert cloned in bacteria has been shown by *in situ* hybridisation to be homologous with the mouse gene and by direct DNA sequence analysis to encode the amino acid sequence of mouse DHFR, (2) DHFR enzymatic activity and resistance to Tp are specified by nucleotide sequences present on chimaeric plasmids but not on the vector, (3) the DHFR

Fig. 6 Filter affinity transfer analysis[29] of bacterial cell extracts. 20 μl of extracts in SDS sample buffer were run at constant current for 3 h in an 11.25% SDS–polyacrylamide slab gel. The gel was incubated in PBS (50 mM phosphate buffer containing 0.15 M NaCl) for 30 min and placed on a blotter wet with PBS. Peptides were specifically transferred from the gel to strips of a dry cellulose filter that had been covalently coupled to anti-DHFR F(ab)$_2$ fragments[12]. Filters were washed, incubated with antibody to DHFR, washed again and treated with ^{125}I-labelled protein A. After additional washing and drying steps, the filters were analysed by autoradiography. The eluate fraction (lane *d*) was obtained by passage of an extract of χ2282 cells containing pDHFR7 over a 0.5 ml methotrexate–Sepharose affinity column. The extract was acidified to *p*H 5.8, passed over the column and the bound fraction was eluted with 2 mM folic acid in a 5 mM $NaHCO_3$ buffer at *p*H 8.5 as described elsewhere[28]. Lane *a*, extracts of mouse cell line; lanes *b* and *c*, pDHFR 7; lane *d*, eluate from methotrexate column; lane *e*, pDHFR21; lane *f*, pBR322.



Nature Vol. 275 19 October 1978

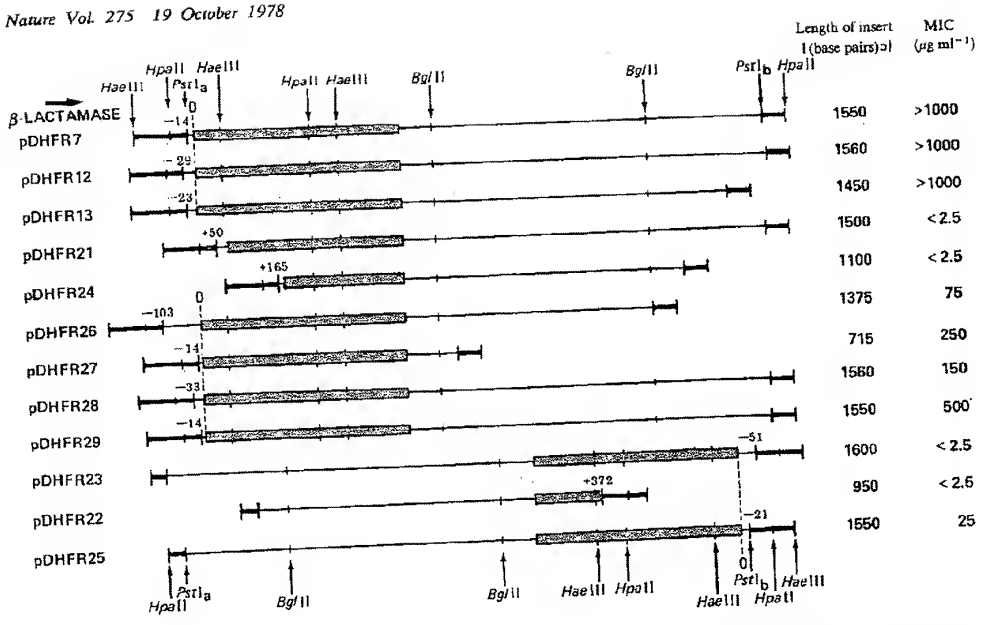


Fig. 7. Some structural and functional properties of chimaeric plasmids and χ2282 clones containing DHFR cDNA insert. The MIC of Tp for each of the χ clones was tested on M9 minimal agar plates containing biotin (2 μg ml^{-1}), casamino acids (0.5%), DAP (50 μg ml^{-1}) and Tp concentrations ranging from 0 to 1,000 μg ml^{-1}. χ2282 (pBR322) was used as a control and is sensitive to Tp at 2.5 μg ml^{-1}, as determined by incubation at 32 °C for 3 d. Plasmid DNA isolated from each clone was digested with *Pst*I endonuclease[15] at 37 °C for 3 h and extracted sequentially with phenol and ether. DNA was precipitated with ethanol, resuspended in 8 μl TE buffer and electrophoresed in 1.2% agarose gels in TBE buffer to determine the length of the inserted fragment. ColEI plasmid DNA digested with *Hae*II endonuclease[42] was added to each sample as an internal molecular weight standard. Additionally, *Hin*dIII-generated fragments of SV40 DNA were used as an external standard[45]. The standard error is ±100 base pairs with the method used. The orientation of the cDNA inserts in the vector plasmid was determined by gel analysis of plasmid DNA digested with *Bgl*I and *Hin*cII endonucleases; in addition, pDHFR25 was treated with the *Eco*RI and *Bgl*I enzymes to confirm the orientation of its insert. The direction of transcription of the β-lactamase gene of pBR322 is indicated by an arrow. The shaded area in each plasmid map indicates the structural sequence for DHFR. The numbers shown above each *Pst*I$_a$ site indicate the distance (in base pairs) between the cleavage point and the first nucleotide of the DHFR structural sequence as described in Fig. 4. This distance was determined exactly by DNA sequence analysis for plasmids pDHFR 7, 12 and 13 and was estimated for the other plasmids by gel electrophoresis of fragments produced by either the *Hae*III or *Hpa*II endonucleases. For the chimaeric plasmids that we subsequently sequenced, this estimate proved to be accurate within 5 base pairs. *Pst*–*Bgl*I distances were determined for the chimaeric plasmids by gel electrophoresis using *Hpa*II-cleaved pBR322 DNA as an internal standard. Nucleotides in the 5′ direction on the mRNA from position 1 have negative numbers and are estimates obtained from gel analysis for all plasmids except pDHFR7 and 12.

encoded by the constructed plasmids shows differential sensitivity to competitive inhibitors of DHFR characteristic of the mammalian gene product, and (4) the enzyme synthesised by bacteria containing DHFR cDNA is immunologically reactive with antibody made against mouse DHFR.

Note that the clones which express the highest levels of Tp resistance contain an immunologically reactive peptide having a size characteristic of the mammalian DHFR. A peptide of this size could potentially result from proteolytic cleavage of a fused β-lactamase–DHFR protein that was initiated at the β-lactamase ribosomal binding site. This proposal is consistent with the finding of large-sized protein species containing DHFR antigenic sites in extracts from cells that are either Tp^R or Tp^S (Fig. 6). More intriguing, however, is the possibility noted above that the *Pst*I–poly dG sequence constructed at the vector–cDNA junction can act together with the nearby ATG (AUG) translational start codon to bind mRNA to the bacterial ribosome and initiate DHFR peptide chains within the cDNA insert. If this interpretation is correct, initiation of peptide chains within other eukaryotic cDNA inserts may be obtainable in bacteria by use of the same structural relationships that have resulted in expression of the mouse DHFR coding sequence. Additional DNA sequence analysis and investigation of the protein products encoded by chimaeric plasmids should provide definitive information on this point and should help elucidate further the structural basis for the different levels of Tp resistance expressed by various clones.

Some of the bacterial clones we have isolated produce proteins that react immunologically with antibody to mouse DHFR but which are not biologically active. Our results suggest that an important obstacle to functional expression of mammalian DNA sequences in bacteria has been the development of an assay capable of detecting those clones that possess both a complete coding sequence and the correct nucleotide relationships to allow such expression. The strong phenotypic selection possible in the present experiments has provided an effective means of identifying and isolating expressing clones.

As the cloned coding sequence for mouse DHFR is selectable in higher organisms as well as in bacteria, it constitutes a powerful tool for the construction of eukaryotic cloning vectors, for the isolation of replication regions of eukaryotic chromosomal and extrachromosomal genomes (compare ref. 30), and for the isolation and characterisation of signals that control genetic transcription and translation in variety of species.

This work was supported by grants from the NIH, American Cancer Society and NSF to S.N.C. and R.T.S. H.A.E. is a Fellow of the American Cancer Society, California Division and acknowledges the hospitality and support of H. O. McDevitt. AMV reverse transcriptase was provided by Dr J. Beard, RNA by Drs F. Alt and R. Kellems, triazine derivative by Dr H. Schaller. TdT was prepared by Dr R. L. Ratliff.

Received 4 September; accepted 15 September 1978.

1. Murrow, J. F. *et al. Proc. natn Acad. Sci. U.S.A.* **71**, 1743–1747 (1974).
2. Struhl, K., Cameron, J. R. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1471–1475 (1976).
3. Ratzkin, G. & Carbon, J. *Proc. natn Acad. Sci. U.S.A.* **74**, 487–491 (1977).
4. Vapnek, D., Hautala, J. A., Jacobson, J. W., Giles, N. H. & Kushner, S. *Proc. nam. Acad. Sci. U.S.A.* **74**, 3508–3512 (1977).
5. Itakura, K. *et al. Science* **198**, 1056–1063 (1977).
6. Villa-Kamaroff, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3727–3731 (1978).
7. Burchall, J. J. & Hitching, G. H. *Molec. Pharmacol.* **1**, 126–136 (1965).
8. *Guidelines for Research Involving Recombinant DNA Molecules*, 23 June, (NIH, US Department of Health, Education and Welfare, 1976).
9. Curtiss, R., III *et al* in *Molecular Cloning of Recombinant DNA*, (eds Scott, W. A. & Werner, R.) 99–111 (Academic, New York, 1977)
10. Miller, J. H. in *Experiments in Molecular Genetics*, 218–220 (Cold Spring Harbor Laboratories, New York, 1972).
11. Grunstein, M. & Hogness, D. S. *Proc. natn. Acad Sci. U.S.A.* **72**, 3961–3965 (1975).
12. Erlich, H. A., Cohen, S. N. & McDevitt, H. O. *Cell* **13**, 681–689 (1978).
13. Alt, F. W., Kellems, R. E., Bertino, J. R. & Schimke, R. T. *J. biol. Chem.* **253**, 1357–1370 (1978)
14. Buell, G. N., Wickens, M. P., Payvar, F. & Schimke, R. T. *J. biol. Chem.* **253**, 2471–2482 (1978).
15. Smith, H., Blattner, N. & Davies, J. *Nucleic Acids Res.* **3**, 343–353 (1976).
16. Kleinschmidt, A. K., Lang, D., Jacherts, D. & Zahn, R. K. *Biochim. biophys Acta* **61**, 856–864 (1962).
17. Bolivar, F. *et al. Gene* **2**, 95–113 (1977)
18. Chang, A. C. Y. & Cohen, S. N. *J Bact.* **134**, 1141–1156 (1978).
19. Enea, V., Vovis, G. F. & Zinder, N. D. *J. molec. Biol.* **96**, 495–509 (1975).
20. Cohen, S. N., Chang, A. C. Y. & Hsu, L. *Proc. natn. Acad. Sci. U.S.A.* **69**, 2110–2114 (1973).
21. Stone, D. & Phillips, A. W. *FEBS Lett.* **74**, 85–87 (1977).
22. Steitz, J. A. & Steege, D. A. *J. molec. Biol.* **114**, 545–558 (1977).
23. Steege, D. A. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4163–4167 (1977).
24. Shine, J. & Dalgarno, L. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1342–1346 (1974).
25. Burchall, J. J. *Ann. N. Y. Acad. Sci.* **186**, 143–152 (1971).
26. Baccanari, D., Phillips, A., Smiles, S., Simky, D. & Burchall, J. J. *Biochemistry* **14**, 5267–5263 (1975).
27. Werkheiser, W. *J. biol. Chem.* **236**, 888–893 (1961).
28. Alt, F. W., Kellems, R. E. & Schimke, R. T. *J. biol. Chem.* **253**, 3063–3074 (1976).
29. Erlich, H. A., Levinson, J., Cohen, S. N. & McDevitt, H. O., in preparation.
30. Tiemais, K., Cabello, F. & Cohen, S. N. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2242–2246 (1975).
31. Seeburg, P. H., Shine, J., Martial, J. A., Baxter, J. D. & Goodman, H. M. *Nature* **270**, 486–494 (1977).
32. Wickens, M. P., Buell, G. N. & Schimke, R. T. *J. biol. Chem.* **253**, 2483–2495 (1978).
33. Chang, L. M. S. & Bollum, F. J. *J. biol. Chem* **246**, 909–916 (1971).
34. Wahl, G. M., Padgett, R. A. & Stark, G. R., in preparation.
35. Roychoudhury, R., Jay, E. & Wu, R. *Nucleic Acids Res.* **3**, 101–116 (1976).
36. Dugaiczyk, A., Boyer, H. W. & Goodman, H. M. *J. molec. Biol.* **96**, 171–184 (1975).
37. McDonell, M. W., Simon, M. N. & Studier, F. W. *J. molec. Biol.* **110**, 119–146 (1977).
38. Kuperstock. Y. M. & Helinski, D. R. *Biochem. biophys. Res. Commun.* **54**, 1451–1459 (1973).
39. Sharp, P. A., Sugden, B. & Sambrook, J. *Biochemistry* **12**, 3055–3063 (1974).
40. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
41. Sanger, F. & Coulson, A. R. *FEBS Lett.* **87**, 107–110 (1978).
42. Kaufman, R. J., Dana, S. L. & Schimke, R. T., in preparation.
43. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. *J biol. Chem.* **193**, 265–275 (1951).
44. Oka, A. & Takanami, M. *Nature* **264**, 193–196 (1976)
45. Fiers, W. *et al. Nature* **273**, 113–120 (1978).

letters to nature

N Galaxies—a new class of X-ray sources

BURBIDGE[1], classified extragalactic objects with bright nuclei into three classes: in order of increasing optical luminosity these are (1) Seyfert galaxies; (2) N galaxies; (3) quasars. Many Seyferts and several quasars have been shown to be X-ray sources. Here we show that N galaxies are also powerful X-ray sources. In fact, all six N galaxies in the 3C radio catalogue with redshifts less than 0.06 (ref. 2) are detected by the first full-sky survey of the Goddard Space Flight Center detectors (A2) on HEAO 1. X-ray emission has also been discovered from a strong Southern Hemisphere radio source, the N galaxy Pic A. Six of the seven objects are classified as broad-line radio galaxies (BLRGs)[3–5] and one (3C371) is classified as a BL Lac object[6]. These objects are differentiated from the newly discovered class of 'emission line' galaxies[7] by their high radio flux and extremely broad lines. Before HEAO 1 there was one confirmed N galaxy, 3C120 (ref. 8), and one suggested, 3C390.3 (refs 9, 10). We detect 3C120, strengthen the identification of 3C390.3, and present evidence for X-ray emission from four new sources, 3C111, 3C382, 3C371, and Pic A. We also suggest 3C445 as the identification of 2A2220-022 (ref. 11).

Figure 1 shows 90% confidence error boxes on the positions of these seven galaxies together with 4U (ref. 10) or 2A (ref. 11) error boxes where applicable. The 2–10 keV luminosity of these objects, as well as other relevant information, is given in Table 1. We have also looked for X-ray flux from the next closest N galaxies, PKS0521 − 36 and 3C227. No significant flux was seen, but the 90% upper limit on the X-ray luminosities (given in Table 2) are consistent with their luminosities being comparable to the seven detected N galaxies. Except for quasars, these N galaxies are the most luminous class of compact X-ray sources yet detected.

Grandi and Osterbrock have classified radio galaxies spectroscopically as BLRG or narrow-line radio galaxies (NLRG) and have demonstrated a high degree of association between the classification 'broad-line' and the morphological type 'N'. In marked contrast to the situation for BLRGs, only 3C405 (Cyg A) of the five NLRGs listed by Grandi and Osterbrock, with redshifts less than 0.06, is a detected X-ray source. 3C317 is confused with Ab2052, a distance class 3 cluster, and so its luminosity is poorly determined. Upper limits on the flux from the other three NLRGs are given in Table 2. The deduced upper limits on the luminosities are considerably less than average X-ray luminosities for BLRG of $\sim 3 \times 10^{44}$ erg s^{-1}. Note that of the detected N galaxies, the one with the narrowest emission lines, 3C371, has the smallest X-ray luminosity.

The lack of X-ray emission from NLRGs contrasted to the virtual certainty of X-ray emission from BLRGs is similar to the distinction between type 1 and type 2 Seyferts, that is, none of the ~20 known X-ray emitting Seyfert galaxies are classical type 2 Seyferts. Grandi and Osterbrock have drawn attention to the optical spectral similarities and differences between BLRGs and type 1 Seyferts, and NLRGs and type 2 Seyferts, respectively.

The fact that many of these N galaxies possess compact radio components suggests that the synchrotron self-Compton process (SSC)[12] may be the principle cause of the X-ray emission. For 3C390.3, the best studied of these objects, a SSC model predicts a physical size to the X-ray emitting region of ~0.1 pc and a time scale for X-ray variability of months[13]. The detection of 3C390.3 by Uhuru and Copernicus at considerably higher flux levels and an Ariel 5 upper limit on 3C382 (ref. 14) ~20% lower than the present detection indicate possible variability in

TAB N

Proc. Natl. Acad. Sci. USA
Vol. 75, No. 12, pp. 5936–5940, December 1978
Biochemistry

Chicken ovalbumin is synthesized and secreted by *Escherichia coli*

(recombinant DNA/gene expression/*lac* control/plasmids)

THOMAS H. FRASER AND BARBARA J. BRUCE

Research Laboratories, The Upjohn Company, Kalamazoo, Michigan 49001

Communicated by Alexander Rich, September 22, 1978

ABSTRACT **By recombinant DNA methods, the chicken ovalbumin structural gene has been fused to *Escherichia coli lac* transcriptional and translational control regions. When a plasmid containing the hybrid gene was introduced into *E. coli*, a protein identified as ovalbumin by immunoreactivity and sodium dodecyl sulfate/polyacrylamide gel electrophoresis was synthesized. The chicken ovalbumin made in bacteria was full length (43,000 daltons) and constituted 1.5% of the cellular protein. In addition, the microbially synthesized ovalbumin was secreted through the cell membrane into the periplasmic space of *E. coli*. The ability of the *E. coli* secretory apparatus to recognize chicken ovalbumin, which is normally synthesized and secreted in hen oviducts, suggests that common features exist in the secretion-recognition mechanisms found in these two organisms. The bacterial synthesis of significant amounts of chicken ovalbumin demonstrates that the *E. coli* cellular machinery may be utilized to synthesize a higher eukaryotic protein which is relatively stable in the bacterial intracellular environment.**

Many of the potential benefits envisioned as a result of the application of recombinant DNA technology to the solution of medical problems require the insertion, into microorganisms, of genes coding for proteins normally found in higher eukaryotic organisms. Although it has been possible to clone a number of different higher eukaryotic genes containing the information necessary to code for proteins, reports of the expression of these proteins in bacteria have been limited to the human polypeptide hormone somatostatin (1) and rat proinsulin (2).

The somatostatin gene was chemically synthesized and the DNA sequence coding for its 14 amino acids was fused to the β-galactosidase structural gene on a plasmid. Yields of somatostatin varied from 0.001 to 0.03% of the total cellular protein. The rat proinsulin gene was inserted into a plasmid-borne penicillinase gene and the bacteria produced approximately 100 molecules of proinsulin per cell.

We report here the *in vitro* construction of a multicopy plasmid in which the *lac* operon control region was fused to a DNA sequence coding for chicken ovalbumin, a 43,000-dalton protein. This egg white protein is normally synthesized and secreted in the chicken oviduct. Bacteria containing the fused plasmid synthesized about 1.5% of their protein as full-length immunoreactive ovalbumin. The production of ovalbumin has been shown to be under *lac* control and the protein is stable *in vitro* to *Escherichia coli* proteases. Because *E. coli* is capable of secreting specific proteins, we also performed experiments to determine whether the *E. coli* secretory machinery can recognize the ovalbumin molecule and found that ovalbumin is secreted through the cell membrane into the periplasmic space.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Cells and DNA. *E. coli* HB101 (UC 6479) (3) was used in all experiments. The *lac* UV5 promotor plasmid pOP203 was constructed by F. Fuller, L. Johnsrud, and W. Gilbert (personal communication) by cloning a 203-base-pair *Hae* III fragment of *lac* DNA containing the promotor, ribosome binding site, and codons for the first seven amino acids of β-galactosidase into the tetracycline resistance plasmid pMB9 (4). The plasmid has a single *Eco*RI site located so that the *lac* promotor will direct transcription into any DNA fragment ligated into that site. The pOV230 plasmid was constructed by L. A. McReynolds, J. F. Catterall, and B. W. O'Malley (5) and was obtained in a cooperative laboratory arrangement. Plasmid DNA was prepared by the salt precipitation method of Guerry *et al.* (6) and further purified by cesium chloride/ethidium bromide density gradient centrifugation (7). Transformation was done essentially as described by Wensink *et al.* (8). Cell extracts were prepared by grinding with alumina, and protein concentrations were determined by the method of Lowry *et al.* (9) with bovine serum albumin as the standard.

Enzyme Reactions. All restriction endonucleases were purchased from New England BioLabs. *Eco*RI, Taq I, and *Hha* I digestions were done in reaction mixtures recommended by the supplier, except that autoclaved gelatin was substituted for bovine serum albumin. Alkaline phosphatase treatment of *Eco*RI-cut pOP203 DNA was carried out essentially as described by Ullrich *et al.* (10), except that the enzyme was preincubated at 70°C for 10 min prior to addition of DNA. T4 DNA polymerase-catalyzed fill-in of staggered ends was done in a reaction mixture described by Goulian *et al.* (11), containing 0.2 mM deoxynucleotide triphosphates.

*Eco*RI linkers, obtained from Collaborative Research, were phosphorylated by using polynucleotide kinase (12) and then heated to 70°C and slowly cooled to 4°C. The reaction mixture was then brought to 45 mM Tris·HCl (pH 7.8), 10 mM 2-mercaptoethanol, 10 mM $MgCl_2$, 15 mM dithiothreitol, 1 mM ATP, 20 μg of *Taq* I-cut pOV230 DNA per ml, and 14 units of T4 DNA ligase (New England BioLabs) per ml and incubated at 12.5°C overnight. Joining of the purified ovalbumin gene fragment to the alkaline phosphatase-treated pOP203 was also catalyzed by T4 DNA ligase (12).

Gel Electrophoresis. DNA was extracted from agarose gels by maceration and overnight incubation at 47°C in 10 mM Tris·HCl, pH 8.0/2 mM EDTA/1 M NaCl. The agarose was pelleted at 100,000 × *g* for 1 hr and the supernatant was phenol extracted, ether extracted, and ethanol precipitated.

Sodium dodecyl sulfate ($NaDodSO_4$)/polyacrylamide gel electrophoresis for protein separation was done as described by Laemmli (13) with a Bio-Rad slab gel apparatus. Samples were dissolved in an equal volume of sample buffer [61.5 mM Tris·HCl, pH 7.8/3% (wt/vol) $NaDodSO_4$/5% (vol/vol) 2-mer-

Abbreviation: $NaDodSO_4$, sodium dodecyl sulfate.

captoethanol/20% (vol/vol) glycerol] and heated at 100°C for 5 min (14).

Immunoassays. The *in situ* immunoassay was done as described by Skalka and Shapiro (15), with the N-Z bottom agar containing 25 μl of ovalbumin antisera (Antibodies, Inc., Davis, CA) per ml, 10 μg of tetracycline per ml, and 1 mM isopropylthiogalactoside.

Immunoreactive material in cell extracts was precipitated for gel analysis by adding antiserum to extracts. The mixtures were allowed to stand for 60 min at room temperature and then overnight at 4°C; the precipitates were washed twice with 25 mM Tris·HCl, pH 7.4/137 mM NaCl/5 mM KCl.

Containment. All recombinant DNA experiments were done under P2-EK1 containment.

RESULTS

Construction of the *lac*–ovalbumin fused plasmid, pUC1001, proceeded as shown in Fig. 1. In our construction we cut the pOP203 plasmid with *Eco*RI and treated the resulting linear molecule with alkaline phosphatase, removing the 5′-phosphate groups to prevent recircularization of the pOP203 vector DNA during ligation (10).

The pOV230 plasmid was constructed with a pMB9 vector and contained nearly all of the ovalbumin mRNA sequence, including all of the information required to code for the amino acid sequence of chicken ovalbumin. The sequence of the ovalbumin gene insert in pOV230 (16) revealed a unique *Taq* I restriction endonuclease site 25 base pairs to the 5′ side of the ovalbumin initiator AUG. Fusion of bacterial control regions at this site would allow expression of the entire ovalbumin structural gene sequence. As shown in Fig. 1, an additional *Taq* I site is located approximately 250 base pairs outside of the ovalbumin insert such that *Taq* I digestion yields a DNA fragment of about 2200 base pairs containing the entire ovalbumin structural gene.

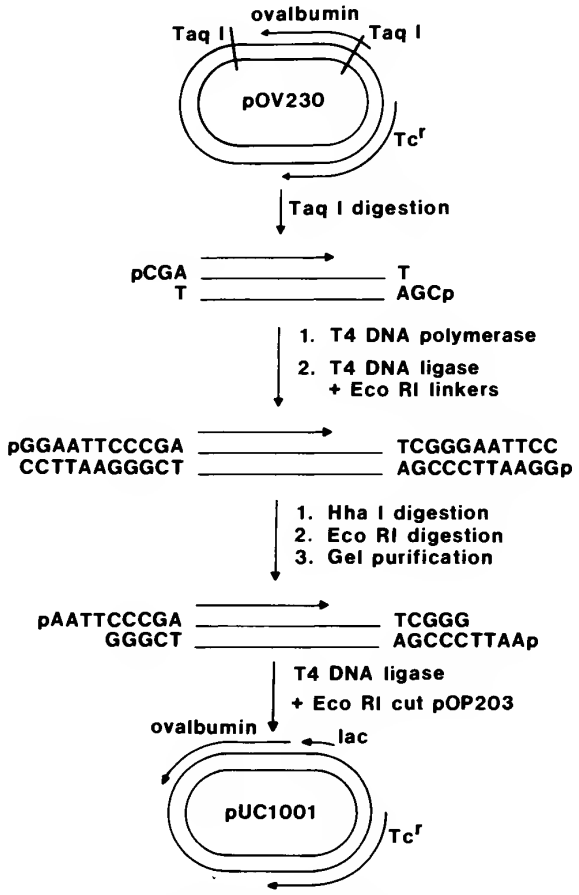


FIG. 1. Outline of steps used to fuse *lac* control elements to the ovalbumin gene, resulting in plasmid pUC1001.

Because the sequences of the *lac* control region inserted into pOP203 (17, 18) and of the ovalbumin gene insert in pOV230 are known, our strategy was to fuse the *lac* and ovalbumin DNAs in a manner that would maintain the translational reading frame, assuming protein synthesis began at the β-galactosidase initiation codon. There are several ways in which this can be done, but we chose to fill in the stagger-ended *Taq* I fragments with DNA polymerase and then ligate synthetic *Eco*RI octamer linkers to the blunt ends as shown in Fig. 1. A large excess of linker molecules was used in the reaction to prevent ligation of the filled-in *Taq* I fragments to each other. The linked fragments were digested with *Eco*RI to generate 5′ staggered ends. Gel purification was facilitated by digestion with *Hha* I to cut a pOV230 fragment that otherwise migrated with the ovalbumin gene.

After elution from the gel, the purified pOV230 DNA fragment was ligated to the alkaline phosphatase-treated pOP203 plasmid vector. This ligated DNA was then used to transform *E. coli* HB101, and tetracycline-resistant colonies were selected. The transformants were assayed for ovalbumin production by an *in situ* immunoassay. Colonies were lysed with lysozyme and Sarkosyl after growth on agar containing ovalbumin antiserum, isopropylthiogalactoside, and tetracycline. Easily discernible precipitin rings formed around several of the colonies, indicating that they were producing substantial amounts of ovalbumin.

In order to verify the key features of the plasmid DNA structure in the ovalbumin-producing bacteria, plasmid DNA from eight transformants was subjected to restriction digestion with *Hae* III alone and with both *Hae* III and *Eco*RI and analyzed by polyacrylamide gel electrophoresis. Fig. 2 illustrates the expected *Hae* III and *Eco*RI cleavage sites in our constructed plasmids. Construction *A* would be expected to synthesize ovalbumin whereas construction *B* would not be expected to synthesize it. *Hae* III digestion would allow determination of the orientation in which the ovalbumin gene is inserted relative to the *lac* control region; *Hae* III/*Eco*RI double digestion would show whether both *Eco*RI sites are present.

Fig. 3 shows the results of these restriction digestion analyses for eight transformants. Lanes 17 and 18 contained the *Hae* III digest and the *Hae* III/*Eco*RI double digest of pOP203. The largest fragment in lane 17 contained the *lac* control region and was 1052 base pairs long. This fragment was cut by *Eco*RI to yield the two fragments, 350 and 700 base pairs, seen in lane 18.

Lanes 1, 3, 5, 7, and 9 contained plasmid DNAs in orientation *A* (Fig. 2); lane 11 contained plasmid DNA in orientation *B*. As shown in lanes 2, 4, 6, 8, 10, and 12, these transformants all contained two *Eco*RI sites.

Lane 13 contained a plasmid in which a deletion apparently occurred and could be explained if the larger band in a plasmid with the ovalbumin gene in orientation *A* had lost some DNA, giving rise to the observed doublet at about 1600 base pairs. The double digest in lane 14 indicates that one of the *Eco*RI sites had been lost, suggesting that a deletion in the 1800-base-pair fragment included the *Eco*RI site. Lanes 15 and 16 contained a plasmid in which a deletion occurred in the pOP203 vector.

The *in situ* immunoassay results showed that colonies con-

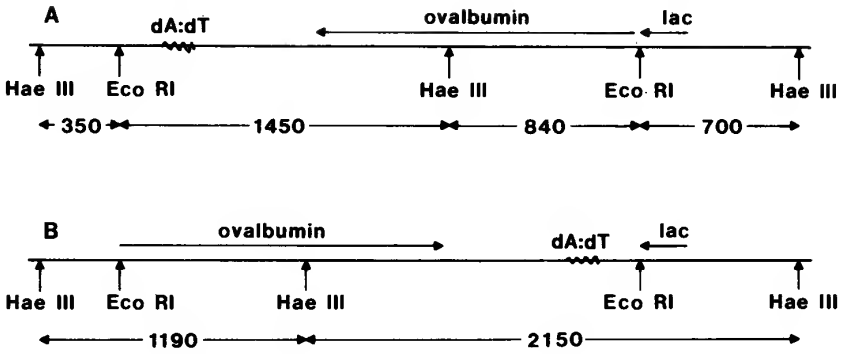


FIG. 2. Restriction digestion maps of plasmids containing the ovalbumin gene, showing the two possible orientations of that gene relative to the *lac* control elements.

taining the plasmids whose *Hae* III digestion patterns are shown in lanes 1, 5, 7, 9, and 13 of Fig. 3 produced ovalbumin. All of these plasmids had the ovalbumin gene inserted in orientation A. The plasmid in lane 3 also was in orientation A but did not produce ovalbumin, perhaps due to a frame shift caused by the addition or deletion of bases during the biochemical manipulations required for fusion.

Fig. 4 is a photograph of a Coomassie blue-stained NaDodSO$_4$/polyacrylamide gel in which several standards as well as immunoprecipitated ovalbumin had been electrophoresed. Lane 2 contained an anti-ovalbumin immunoprecipitate of an *E. coli* HB101 (pUC1001) extract. The top band is immunoglobulin heavy chain (50,000 daltons); the bottom band is immunoglobulin light chain (23,500 daltons). The dark band between the heavy and light chains in lane 2 is the *E. coli* ovalbumin. Although several other very faint bands are visible, it is clear that the immunoprecipitation is specific.

Lane 1 in Fig. 4 contained an immunoprecipitated *E. coli* HB101 (pUC1001) extract that had been subjected to incubation for 4 hr at 37°C. The size of the *E. coli* ovalbumin was not reduced by this self-digestion. Lanes 3, 4, and 5 contained protein standards to assist in the determination of the molecular weight of *E. coli* ovalbumin. Lane 3 contained alkaline phosphatase (43,000 daltons), lane 4 contained purified egg white ovalbumin, and lane 5 contained periodate-treated (19) egg white ovalbumin. Whereas the untreated egg ovalbumin migrated slightly faster than the *E. coli* ovalbumin, gentle treatment with periodate, which should remove carbohydrate residues, caused them to migrate at the same rate.

By using a gel diffusion immunoassay with egg white ovalbumin as a standard it was found that between 1.35 and 1.6% of the cellular protein was ovalbumin. Spectrophotometric scanning of Coomassie blue-stained NaDodSO$_4$/polyacrylamide gels also indicated that about 1% of the cellular protein was ovalbumin.

We found about a 50% decrease in the amount of ovalbumin produced in the absence of isopropylthiogalactoside, indicating that synthesis is under *lac* control. We expected only a small decrease because the multicopy pUC1001 plasmids contain functional *lac* operator, resulting in nearly constitutive levels of transcription in the absence of inducer.

To determine whether *E. coli* secreted ovalbumin, bacteria containing the pUC1001 plasmid were grown in the presence of isopropylthiogalactoside and tetracycline. The cells were treated with lysozyme and EDTA to give protoplasts. Centrifugation of the protoplast preparation yielded a supernatant fraction containing periplasmic proteins and a pellet containing protoplasts. β-Galactosidase was used as a marker for nonsecreted proteins and alkaline phosphatase was used as a marker for periplasmic proteins.

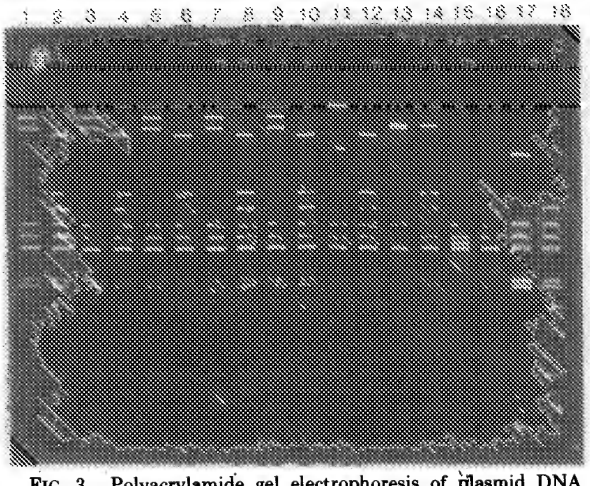


FIG. 3. Polyacrylamide gel electrophoresis of plasmid DNA restriction digestion fragments. The odd-numbered lanes are *Hae* III plasmid digestions and the even-numbered (n + 1) lanes are the same plasmid subjected to *Hae* III/*Eco*RI double digestion.

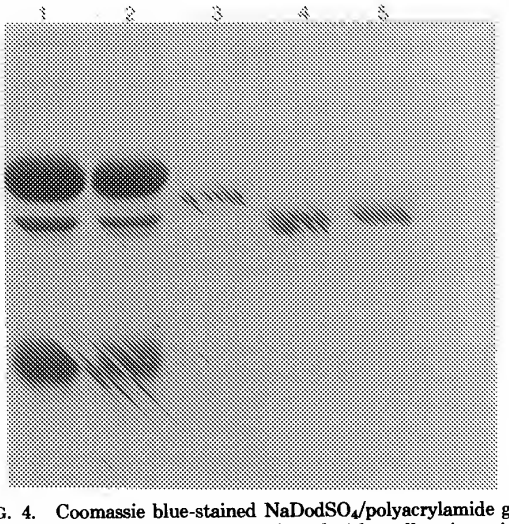


FIG. 4. Coomassie blue-stained NaDodSO$_4$/polyacrylamide gel. Lanes: 1 and 2, *E. coli* extracts precipitated with ovalbumin antiserum; 3, alkaline phosphatase; 4, purified egg white ovalbumin; 5, periodate-treated egg white ovalbumin.

Table 1. Results of protoplasting experiment

Fraction	β-Galactosidase activity		Alkaline phosphatase activity		Ovalbumin immunoreactivity	
	Units	% total	Units	% total	μg	% total
Periplasmic	1,005	3.7	32.24	73.4	620	48.6
Protoplast	26,189	96.3	11.7	26.6	656	51.4

Protoplasts of HB101 (pUC1001) were prepared essentially as described by Malamy and Horecker (20). β-Galactosidase and alkaline phosphatase were assayed as described (21, 22). Units are defined as the enzyme activities required to produce an absorbance change of 1.0/hr; the units shown represent the total activities from a 600-ml broth culture.

Table 1 shows the results of a protoplasting experiment. More than 96% of the β-galactosidase activity was in the protoplast fraction, indicating less than 4% leakage of this nonsecreted protein into the periplasmic fraction. Less than 75% of the alkaline phosphatase activity was in the periplasmic fraction, suggesting that some of the periplasmic protein is not released into the supernatant. Table 1 shows that approximately 50% of the ovalbumin was found in the periplasmic fraction and 50% in the protoplast fraction. In view of the extremely low level of β-galactosidase leakage into the periplasmic fraction, this indicates that the chicken ovalbumin was actively secreted by *E. coli*, although perhaps not as efficiently as *E. coli* secretory proteins such as alkaline phosphatase. In no case was any immunoreactive ovalbumin found in the growth medium. NaDodSO$_4$/polyacrylamide gel electrophoresis of anti-ovalbumin immunoprecipitated periplasmic and protoplast fractions has shown that the secreted and nonsecreted ovalbumin have the same molecular weight.

DISCUSSION

We have shown that full-length ovalbumin molecules are synthesized in bacteria from information encoded in the pUC1001 plasmid. Because the *E. coli* ovalbumin expression from pUC1001 is induced by isopropylthiogalactoside, we believe that ovalbumin mRNA synthesis is initiated at the plasmid *lac* promotor. From Fig. 5, showing the first 107 bases of this mRNA, it can be seen that if protein synthesis is initiated at the β-galactosidase AUG (position 39) the ovalbumin molecule will be synthesized with 18 extra amino acids on its amino terminus. Addition or deletion of bases during the *in vitro* construction of pUC1001 would result in a frame shift, and ovalbumin would not be expressed. This seems to be the most likely explanation for lack of expression from the plasmid in lane 3 of Fig. 3 and, if true, would demonstrate that protein synthesis is initiated at the β-galactosidase AUG.

Our data, however, do not allow us to conclude directly whether the β-galactosidase protein synthesis initiation site is used for ovalbumin expression. The results from polyacrylamide gel electrophoresis show that the *E. coli* ovalbumin has a molecular weight of about 43,000, but this value is dependent upon standards whose gel migration characteristics have not been extensively characterized. It is possible that the 18-amino acid tail is synthesized and then cleaved from the molecule by intracellular proteases. This question will have to be looked at in more detail, however, before any firm conclusions may be drawn.

Although 1.5% of the cellular protein is ovalbumin in *E. coli* carrying the pUC1001 plasmid, this level is considerably lower than the level theoretically attainable in this system. We have determined (unpublished data) that there are between 25 and 30 copies of the pUC1001 plasmid per chromosome equivalent in *E. coli*. It is known that β-galactosidase, with one gene copy per chromosome equivalent, is expressed as 2% of the total *E. coli* protein in fully induced cultures (23). Therefore, considering the gene copy numbers as well as the differences in molecular weight between ovalbumin and β-galactosidase, we would expect approximately 19% of the total protein in our pUC1001-containing bacteria to be ovalbumin. In fact, although a significant amount of ovalbumin was synthesized, it was less than 10% of the theoretically predicted amount.

The most probable explanations for lower-than-theoretical levels of ovalbumin in cells carrying the pUC1001 plasmid are inefficient transcription, inefficient translation, and proteolytic degradation of ovalbumin molecules.

It has been reported that, when a *lac* promotor with the UV5 (CAP-independent) mutation is cloned onto a multicopy plasmid, it functions with full efficiency (24). Thus, it is also likely that transcription is fully efficient in the case of pUC1001. Inefficient translation of the ovalbumin mRNA may result from either inefficient initiation or inefficient elongation. It is known that at least 7 times the normal level of β-galactosidase can be made when the gene is on a multicopy plasmid (25). Therefore, it is likely that protein synthesis initiation is not limiting in our system. The rate of elongation of mRNA translation will be dependent upon many factors, one of which is the availability of aminoacyl-tRNAs with anticodons complementary to the codons on the mRNA. The sequence data from several prokaryotic and eukaryotic genes, including ovalbumin, have revealed nonrandom utilization of synonymous codons (16, 26, 27). Although not enough data have been collected to permit

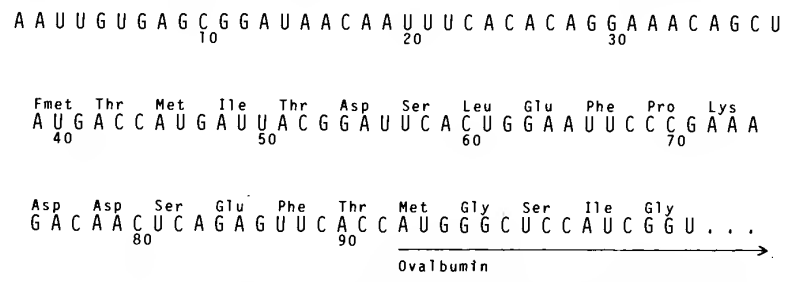


FIG. 5. Predicted pUC1001 mRNA sequence initiated at the *lac* promotor and predicted amino acid sequence initiated at the β-galactosidase AUG.

any generalizations, it is probable that some of the tRNA isoaccepting species required in relatively large amounts for the translation of ovalbumin are present in *E. coli* in low concentrations, thus limiting synthesis.

The third explanation for lower-than-theoretical ovalbumin production is proteolytic degradation. This may occur either with nascent polypeptide chains on polysomes or with mature ovalbumin molecules, or both. Our self-digestion data with *E. coli* cell extracts suggest that the mature ovalbumin molecules are stable to proteolytic digestion *in vitro*, but our data do not extend to the stability of nascent polypeptide chains.

The finding that ovalbumin molecules synthesized in *E. coli* are secreted into the periplasmic space suggests that common features exist in the secretion-recognition mechanisms found in chicken oviducts and in *E. coli*. More experiments will have to be done in order to determine what structural features of the bacterial ovalbumin cause it to be secreted. The ovalbumin naturally secreted from chicken oviducts is a complex protein, containing an amino-terminal acetyl group, carbohydrate residues, and phosphate residues. It is not clear what role, if any, these post-translational modifications play in secretion, and it has not been determined whether or not the bacterial ovalbumin is modified.

Whatever the mechanism for ovalbumin secretion in the hen oviduct, it would now appear that some aspects are common among living organisms. In this regard it should be noted that the secretory "signal" for chicken ovalbumin is probably not a hydrophobic amino-terminal sequence (28), as is the case with many secretory proteins (29, 30, 31). Thus, there may be additional structural features that are generally recognized for secretion.

Although further work is needed to maximize the synthesis of eukaryotic proteins in bacteria, the microbial production of significant amounts of chicken ovalbumin reported here demonstrates that the *E. coli* cellular machinery may be utilized to synthesize a eukaryotic protein that is stable in the bacterial intracellular environment. There is no reason to believe that the synthesis of other high molecular weight animal proteins will be any more difficult than that of ovalbumin.

Note Added in Proof. Results similar to those presented here involving synthesis of an ovalbumin-like protein by *E. coli* harboring a recombinant plasmid have recently been reported (32).

We thank Dr. B. W. O'Malley for pOV230 and for supplying pOV230 sequence data prior to publication. We also thank L. Korn for his computer program, F. Fuller for supplying the pOP203 plasmid, and Dr. V. S. Malik for suggesting it as a source of *lac* control elements.

1. Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F. & Boyer, H. W. (1977) *Science* **198**, 1056–1063.
2. Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3727–3731.
3. Boyer, H. W. & Roulland-Dussoix, D. (1969) *J. Mol. Biol.* **41**, 459–472.
4. Rodriguez, R. L., Bolivar, F., Goodman, H. M., Boyer, H. W. & Betlach, M. C. (1976) in *Molecular Mechanisms in the Control of Gene Expression*, eds. Nierlich, D. P., Rutter, W. J. & Fox, C. F. (Academic, New York), pp. 471–477.
5. McReynolds, L. A., Catterall, J. F. & O'Malley, B. W. (1977) *Gene* **2**, 217–231.
6. Guerry, P., LeBlanc, D. J. & Falkow, S. (1973) *J. Bacteriol.* **116**, 1064–1066.
7. Clewell, D. & Helinski, D. (1970) *Biochemistry* **9**, 4428–4440.
8. Wensink, P. C., Finnegan, D. J., Donelson, J. E. & Hogness, D. S. (1974) *Cell* **3**, 315–325.
9. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. (1951) *J. Biol. Chem.* **193**, 265–275.
10. Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W. J. & Goodman, H. M. (1977) *Science* **196**, 1313–1319.
11. Goulian, M., Lucas, Z. J. & Kornberg, A. (1968) *J. Biol. Chem.* **243**, 627–638.
12. Sugino, A., Goodman, H. M., Heyneker, H. L., Shine, J., Boyer, H. W. & Cozzarelli, N. R. (1977) *J. Biol. Chem.* **252**, 3987–3994.
13. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
14. Moir, A. & Brammar, W. J. (1976) *Mol. Gen. Genet.* **149**, 87–99.
15. Skalka, A. & Shapiro, L. (1976) *Gene* **1**, 65–69.
16. McReynolds, L., O'Malley, B. W., Nisbet, A. D., Fothergill, J. E., Givol, D., Fields, S., Robertson, M. & Brownlee, G. G. (1978) *Nature (London)* **273**, 723–728.
17. Maizels, N. M. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 3585–3589.
18. Dickson, R. C., Abelson, J., Barnes, W. M. & Reznikoff, W. S. (1975) *Science* **187**, 27–35.
19. Spiro, R. G. (1966) *Methods Enzymol.* **8**, 44–49.
20. Malamy, M. H. & Horecker, B. L. (1964) *Biochemistry* **3**, 1889–1892.
21. Lederberg, J. (1950) *J. Bacteriol.* **60**, 381–392.
22. Torriani, A. (1966) in *Procedures in Nucleic Acid Research*, eds. Cantoni, G. L. & Davies, D. R. (Harper & Row, New York), pp. 224–235.
23. Miller, J. H. (1972) in *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 398–404.
24. Backman, K. & Ptashne, M. (1978) *Cell* **13**, 67–71.
25. O'Farrell, P. H., Polisky, B. & Gelfand, D. H. (1978) *J. Bacteriol.* **134**, 645–654.
26. Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchison, C. A., III, Slocombe, P. M. & Smith, M. (1977) *Nature (London)* **265**, 687–695.
27. Efstratiadis, A., Kafatos, F. C. & Maniatis, T. (1977) *Cell* **10**, 571–585.
28. Palmiter, R. D., Gagnon, J. & Walsh, K. A. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 94–98.
29. Habener, J. F., Rosenblatt, M., Kemper, B., Kronenberg, H. M., Rich, A. & Potts, J. T., Jr. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2616–2620.
30. Inouye, H. & Beckwith, J. (1978) *Proc. Natl. Acad. Sci. USA* **74**, 1440–1444.
31. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67**, 835–851.
32. Mercereau-Puijalon, O., Royal, A., Cami, B., Garapin, A., Krust, A. Gannon, F. & Kourilsky, P. (1978) *Nature (London)* **275**, 505–510.

TAB O

MOPC-149 fragment is shown in Fig. 5*b*. The fragments are homologous for 13.5 kilobases, but also diverge at their 5′ end. In each case, the divergent heteroduplex bubble represents DNA that is different in the embryonic sequence. Evidently these new sequences are derived from a recombination-like event which joined DNA sequences to the 5′ ends of plasmacytoma DNA fragments and eliminated sequences present in the embryonic fragment (see diagram, Fig. 6). By measuring the lengths of the single-stranded DNAs found in each heteroduplex, we can estimate the lengths of the new DNA sequences that have been joined to the embryonic *Eco*R1 fragment. The points at which the embryonic fragment differs from each rearranged fragment defines the site at which the recombination event took place.

The recombination events that formed the new MOPC-149 and MOPC-41 fragments involved different sites. In MOPC-149 the recombination event appears to have occurred 2.7 kilobases from the constant region gene and removed the MOPC-41-like J region sequence from the embryonic fragment. Presumably the MOPC-149 light chain uses a different J region sequence. The recombination event that generated the MOPC-41 light chain sequence occurred 3.7 kilobases from the constant region gene, very close to the J region sequence found in the embryonic fragment. Here a single recombination event may have linked the MOPC-41 variable region to the MOPC-41 J region to generate the cloned fragment. However, until the nucleotide seuences of the appropriate regions have been determined, we cannot be certain that the recombination event that generated the MOPC-41 gene did not actually replace the embryonic J region with a MOPC-41 J region at the same relative position. It is also possible that both J region sequences are present on the MOPC-41-derived fragment. Nevertheless, these analyses indicate that the recombination events that affect immunoglobulin light chain genes can involve sites at least 1.0 kilobases apart.

The emerging picture of antibody gene arrangement and rearrangement

By characterising cloned, κ chain-encoding segments of DNA derived from mouse embryo and antibody producing cells, we demonstrate directly that immunoglobulin constant and variable region genes undergo a somatic rearrangement that brings together the segments necessary to encode a complete κ light chain gene. Moreover, in each light chain producing plasmacytoma that we have thus far examined there is evidence of somatic rearrangement of one set of these genes, while a second, presumably allelic, set retains the original embryonic pattern. These observations fit well with the plausible notion that rearrangement of a single allele is instrumental in immunoglobulin gene activation and that it may play a part in allelic exclusion in antibody production.

In addition we see that, like the mouse λ light chain[9], κ genes are encoded in at least three segments in embryonic DNA, a foreshortened variable region sequence[4], a short J sequence probably encoding the last 10 or so amino acids of the variable region and, finally, the constant region sequence. Rearrangement of these sequences (which may itself contribute to the generation of antibody diversity) forms, in the case of the MOPC-41 light chain gene, a new chromosomal sequence in which variable and J sequences are adjacent, but separated from the constant region gene by a large 3.7-kilobase intervening sequence of DNA. In the case of MOPC-149, this rearrangement has eliminated the embryonic J sequence and replaced it, in our clone, with a new segment of DNA. Whether this new segment is a portion of the intervening sequence linking MOPC-149 variable and constant regions in genomic DNA is not known. In contrast to what might have been expected, these rearrangements do not occur at a single site, but the new DNA segments are joined at sites about 1.0 kilobases apart on the 5′ side of the constant region gene. The precise nature and the molecular basis of this recombination remains unclear. Further studies should indicate whether or not reciprocal recombination is involved and whether this recombination is mediated by regions of sequence homology.

We thank Dr Fred Blattner for the λCH4A phage system, Barbara Norman, Marion Nau and Aya Leder for help in the gene purification and cloning experiments, Margery Sullivan for help with the electron microscopy and Dr Henry I. Miller for help with preparative gel electrophoresis. We thank Aya Leder for suggesting the gene orientation experiments shown in Fig. 3*b*, and Terri Broderick for assistance in preparation of this manuscript.

Received 17 October; accepted 27 November 1978.

1. Dreyer, W. J. & Bennett, J. C. *Proc. natn. Acad. Sci. U.S.A.* **54**, 864–868 (1965).
2. Tonegawa, S., Hozumi, N., Matthyssens, G. & Schuller, R. *Cold Spring Harb. Symp. quant. Biol.* **41**, 877–889 (1976).
3. Tonegawa, S., Brack, C., Hozumi, N. & Schuller, R. *Proc. natn. Acad. Sci. U.S.A.* **8**, 3518–3522 (1977).
4. Seidman, J. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3881–3885 (1978).
5. Hozumi, N. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3628–3632 (1976).
6. Tonegawa, S., Brack, C., Hozumi, N. & Pirrotta, V. *Cold Spring Harb. Symp. quant. Biol.* **42**, 921–931 (1977).
7. Rabbitts, T. H. & Forster, A. *Cell* **13**, 319–327 (1978).
8. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1485–1489 (1978).
9. Brack, C., Hirama, M., Lenhard-Schuller, R. & Tonegawa, S. *Cell* **15**, 1–14 (1978).
10. Kabat, E. A., Wu, T. T. & Bilofsky, H. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2429–2433 (1978).
11. Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. *Science* **202**, 11–17 (1978).
12. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
13. Tilghman, S. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **10**, 4406–4410 (1977).
14. Tiemeier, D. C. *et al. Cell* **14**, 1309–1313 (1978).
15. Leder, A. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
16. Blattner, F. R. *et al. Science* **196**, 161–169 (1977).
17. Bek, A. & Gould, M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 581–585 (1975).
18. Sternberg, N., Tiemeier, D. & Enquist, L. W. *Gene* **1**, 255–280 (1977).
19. Seidman, J. G., Edgell, M. H. & Leder, P. *Nature* **271**, 582–585 (1978).
20. White, R. L. & Hogness, D. S. *Cell* **10**, 177–192 (1977).
21. Tilghman, S. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 725–729 (1978).
22. Lenhard-Schuller, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 4709–4713 (1978).

Synthesis of growth hormone by bacteria

P. H. Seeburg, J. Shine, J. A. Martial, R. D. Ivarie, J. A. Morris, A. Ullrich, J. D. Baxter & H. M. Goodman

Howard Hughes Medical Institute Laboratory, Department of Biochemistry and Biophysics, Department of Medicine, and Metabolic Research Unit, University of California, San Francisco, California 94143

A hybrid gene was constructed between the β-lactamase gene of plasmid pBR322 and the cloned coding sequence for rat growth hormone. This gene is expressed in bacteria and growth hormone sequences are detectable by immunological methods.

RECENT advances in recombinant DNA technology have led to the cloning in bacteria of the natural coding sequences for several mammalian peptide hormones[1–3]. These sequences provide excellent material for testing the feasibility of producing mammalian proteins in bacteria. The success of such production depends on the correct design and construction of new genes in which suitable bacterial control elements promote expression of

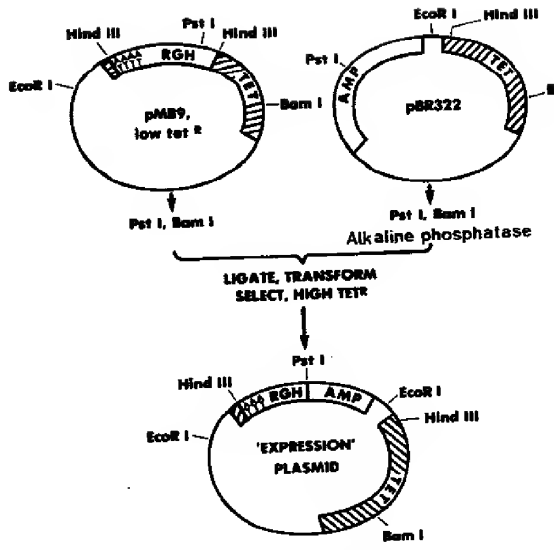


Fig. 1 Construction of a vector for bacterial synthesis of GH sequences. The cloned RGH cDNA was cleaved from plasmid pRGH-1 (ref. 2) with *Hin*dIII endonuclease and isolated by polyacrylamide gel electrophoresis. This DNA was enzymatically ligated to similarly cleaved and alkaline phosphatase-treated[1,3] pMB9. The ligation mixture was used to transform *E. coli* strain χ1776 (refs 1–3). Colonies containing plasmids were selected by growth on low levels (5 μg ml^{-1}) of tetracycline[11]. Resistant colonies were examined for the presence of plasmids containing inserted RGH sequences by gel analysis of such plasmids after digestion with restriction endonuclease *Pst*I and *Eco*RI. In some plasmids RGH DNA was found to have its poly dA · dT end orientated towards the single *Eco*RI site in pMB9. Such a plasmid (pMB9-RGH) was used in the subsequent steps for vector construction. The pre-RGH sequences from pMB9 were placed under the control of the β-lactamase gene of pBR322, as indicated. Plasmid pMB9-RGH was digested with *Pst*I plus *Bam*HI endonuclease. The single cleavage site for *Pst*I in this plasmid occurs between amino acids −24 and −23 of the pre-sequence of RGH (see Fig. 2), whereas the single *Bam*HI site is in the gene responsible for bacterial tetracycline resistance[9]. The amp^r and tet^r plasmid pBR322 was similarly digested with these two enzymes. In pBR322 the *Pst*I site occurs at a sequence coding for amino acids 182 and 183 of pre-β-lactamase[10]. The *Bam*HI site is in the *tet*^{r} gene and is in exactly the same place as in pMB9, as this region of pBR322 was derived from pMB9 (ref. 9). After digestion, pBR322 was treated with bacterial alkaline phosphatase[1,3]. This mixture was then ligated to the *Pst*I and *Bam*HI cleaved pMB9-RGH. After transformation of *E. coli*, colonies were selected for their ability to grow on high levels (20 μg ml^{-1}) of tetracycline.

Table 1 Radioimmunoassay of RGH in extracts from minicells containing pEx-RGH or pBR322

Plasmid	RGH by radioimmunoassay (ng ml^{-1})	
	Spheroplasts	Periplasmic space
pEx-RGH	89.1 (70.8–116.6)	30.8 (30.0, 31.6)
pBR322	Undetectable	Undetectable

One litre of *E. coli* P648-54 containing either pEx-RGH or pBR322 was grown to saturation and collected by centrifugation at 10,000*g* for 10 min. The cells were washed and spheroplasts were prepared by modification of procedures described elsewhere[15]. Briefly, cells were washed once in 200 ml of 10 mM Tris (*p*H 7.4) containing 1 mM PMSF (phenylmethylsulphonyl fluoride, Sigma). The cells were resuspended in 50 ml of 20% sucrose, 0.033 Tris (*p*H 8.0) and 1 mM PMSF. Ice-cold 0.1 M EDTA (*p*H 8.0) was added (0.5 ml) and the mixtures were kept on ice for 5 min. After addition of 1 ml of 5 mg ml^{-1} lysozyme (Sigma) the mixtures were incubated for 30 min at 0 °C and centrifuged (17,000*g*, 10 min, 0 °C). The supernatant media were saved for radioimmunoassay of material from the periplasmic space. The pellets were resuspended in 10 mM Tris (*p*H 8.0) and 1 mM PMSF. $MgCl_2$ was added to a final concentration of 10 mM with 200 μl each of DNase (1 mg ml^{-1}) and RNase (5 mg ml^{-1}). Incubation proceeded for 1 h at 4 °C (spheroplast sample). These mixtures (10 ml) and 50 ml of the samples containing periplasmic material were precipitated with $(NH_4)_2SO_4$ (50% saturation). The samples were resuspended in 4 ml of 50 mM Tris (*p*H 9.0), 0.1 M NaCl and 1 mM PMSF and dialysed against 2 l of the same buffer for 2 h at 4 °C. Serial dilutions of these samples were analysed by radioimmunoassay as described previously[19]. Amounts of RGH found by this assay are listed in ng ml^{-1} with the mean and range of these determinations for the spheroplast sample and the mean and individual values for duplicate samples from the periplasmic space sample. No activity was detected in samples from pBR322. The sensitivity minimum of the assay as assessed from the standard curve with authentic GH was 10 ng ml^{-1}. The high immune reactivity in the periplasmic space is probably due to lysis of some cells during spheroplast formation.

the appropriate coding sequences. We describe here the construction of such a hybrid gene on a bacterial plasmid designed to program bacteria for the synthesis of pituitary growth hormone (GH) sequences. This peptide hormone is crucial for linear growth. A sufficient supply would allow wider therapeutic use in the treatment of growth disorders[4–6] and other conditions in man and might be useful in animal husbandry for more economical food production. The species specificity and size (190 amino acids) of GH[7,8] prevents us from exploiting existing resources or applying current synthetic techniques. Thus, for GH and similarly for a variety of other eukaryotic proteins, genetically programmed bacteria may provide an important resource.

Gene construction

The construction of a hybrid gene between bacterial and mammalian DNA sequences, capable of directing synthesis of rat growth hormone (RGH) sequences in bacteria, is shown in Fig. 1. The bacterial sequences for this gene were excised from the multicopy plasmid pBR322 (ref. 9). They contain a portion of the gene (*amp*^{r}) coding for the enzyme β-lactamase which is secreted into the periplasmic space and is responsible for bacterial resistance to the antibiotic ampicillin[9]. The mammalian sequences were provided by cloned complementary DNA containing the entire coding region for the precursor form of rat pituitary GH[2]. Functional fusion of the two donor DNAs was facilitated by knowledge of their primary structures[2,10]. Both DNAs possess a unique cleavage site for the restriction endonuclease *Pst*I at which they can be joined to maintain the correct reading frame (Fig. 2). In the *amp*^{r} gene the *Pst*I site is situated approximately two-thirds towards the end. In RGH cDNA the corresponding site occurs a few base pairs distal to the

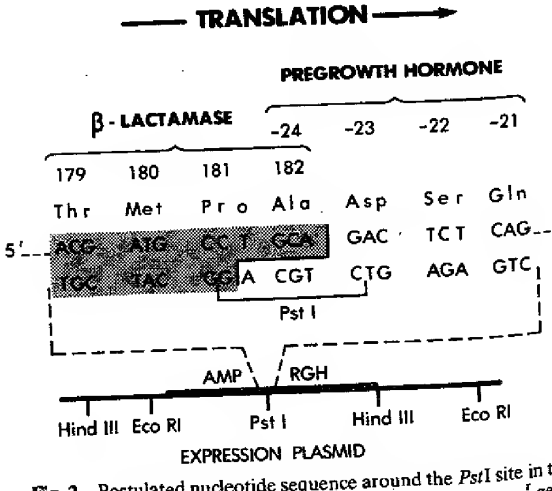


Fig. 2 Postulated nucleotide sequence around the *Pst*I site in the hybrid gene. The *Pst*I site is at the fusion between the *amp*^{r} gene sequences[10] and the cloned RGH cDNA[2]. Positive numbers refer to amino acid residues of pre-β-lactamase and negative numbers to those of the RGH pre-sequence.

initiator codon for the pre-peptide. The resultant new gene should code for a chimaeric protein of 395 amino acids (molecular weight ~44,000) containing the N-terminal 181 amino acids of the β-lactamase precursor covalently linked to 214 residues of rat pre-GH. As this fusion protein carries the signal peptide of β-lactamase, it seems possible that it might be transported to the periplasmic space. Such an extracellular location might minimise possible protease attack and facilitate isolation.

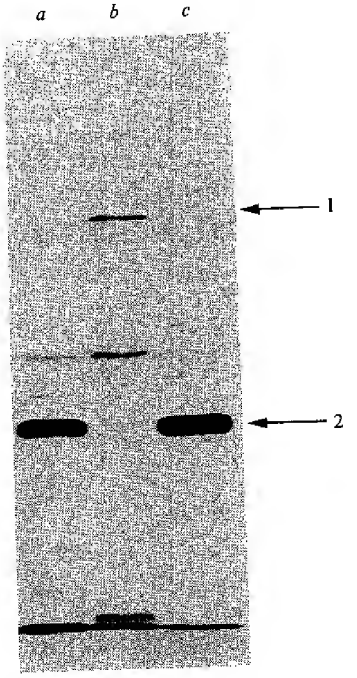


Fig. 3 Proteins synthesised from plasmids pEx-RGH and pBR322. The minicell-producing *E. coli* strain P678-54 (ref. 13) was transformed with pEx-RGH or pBR322 by standard methods[2]. A culture of each (200 ml) was grown to the late log or stationary phase of growth ($A_{650} = 0.7$) in L broth containing 0.5% glucose. Minicells were purified by repeated zone sedimentation through 5–25% sucrose gradients as described elsewhere[12] and resuspended in 2 ml of leucine- and methionine-free M-9 salts medium. After 30 min preincubation at 37 °C, 250 μCi of ^{35}S-methionine and 50 μCi of ^{3}H-leucine were added to both cell preparations, which were then vigorously shaken for 30 min at 37 °C. After a 30-min chase period with 400 μg of nonradioactive methionine and leucine, the minicells were pelleted by centrifugation at 8,000*g* for 2 min and half the cells of each preparation were lysed in 200 μl of SDS sample buffer[14]. The second half of each minicell preparation was used to isolate the proteins secreted into the periplasmic space as described elsewhere[15]. Labelled proteins isolated from equivalent amounts of minicells were electrophoresed on a 10% polyacrylamide–SDS gel[14] and identified by autoradiography of the dried gels. Lanes *a* and *b* represent the total proteins synthesised by pBR322 and pEx-RGH, respectively. Lane *c* reflects the proteins secreted into the periplasmic space of minicells carrying pBR322. Arrow 1 refers to the location of the fusion protein and arrow 2 indicates β-lactamase.

To enable selection of plasmids carrying the correct hybrid gene, the cloned GH cDNA (originally isolated in plasmid pBR322) was first transferred from the *Hin*dIII site of pBR322 into the homologous site of pMB9, a closely related plasmid which lacks the *amp*^{r} gene. In the resultant recombinant plasmid (pMB9-RGH) the gene for tetracycline resistance is expressed at reduced levels[11]. One such plasmid which contains the GH cDNA inserted in the orientation shown in Fig. 1 was used as the recipient of *amp*^{r} gene sequences. Replacement of the unique

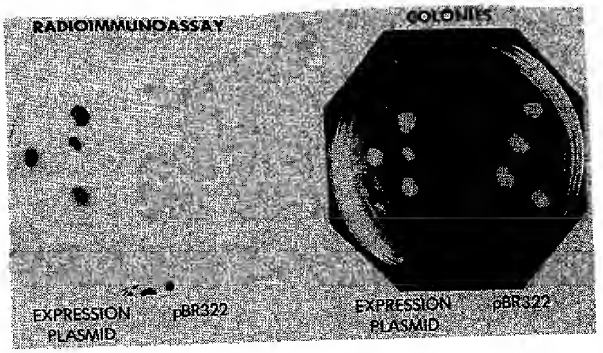


Fig. 4 Immunological detection of RGH sequences in bacteria. Bacterial colonies containing the constructed expression vector pEx-RGH were assayed for the presence of RGH sequences using a modification[16] of the solid-phase immunological screening method Broome and Gilbert[17]. The described wash buffer was replaced with phosphate-buffered saline (0.1 M NaCl, 0.025 M potassium phosphate, *p*H 7.4) containing gelatin (10 mg ml^{-1}) and NP40 (0.1%). Bacterial colonies of *E. coli* strain P 678-54 (ref. 13) carrying pEx-RGH or pBR322 were grown on L-broth agar[18] in Petri dishes. A photograph of such a plate is shown on the right. After treatment (20 min) with chloroform vapour the colonies were exposed to polyvinyl strips coated with antiserum to RGH. Strips were then soaked in the modified wash buffer (see above) to minimise binding of IgG in the subsequent reaction. The strips were then incubated overnight at 4 °C with affinity column-purified ^{125}I-anti-RGH IgG (10^6 c.p.m., specific activity ~5.3 μCi per μg), washed, and exposed to X-ray film. An autoradiogram is shown on the left.

*Pst*I–*Bam*HI fragment in pMB9-RGH with that from pBR322 should place the coding sequences for rat GH under control of the *amp*^{r} gene and will at the same time restore a fully functional *tet*^{r} gene (see Fig. 1). Thus, this fragment replacement enabled selection of bacterial colonies carrying the expression vector 'pEx-RGH' by their ability to grow on high levels of tetracycline.

Gene expression

Purified minicells[12] were used for the study of proteins encoded by the expression vector pEx-RGH. Proteins were analysed by SDS–polyacrylamide gel electrophoresis and compared with those encoded by pBR322 (Fig. 3) and pMB9 (data not shown). As anticipated, the two polypeptides pre-β-lactamase and β-lactamase which are synthesised in cells carrying pBR322 are not made from pEx-RGH. The expression vector codes uniquely for a polypeptide (Fig. 3, lane *b*, arrow 1) with an apparent molecular weight (46,000) within error limits of that expected for the hybrid protein. A comparison of band intensities (corrected for cell number) from several experiments similar to that described in Fig. 3 suggests that the 46,000 MW protein is produced in substantial quantities from pEx-RGH, although in lower amounts (about one-fifth) than β-lactamase from pBR322. When proteins released after the formation of spheroplasts in high osmotic pressure medium[15] were analysed by gel electrophoresis, a large portion of β-lactamase (Fig. 3) but not of the pEx-RGH specific product was detected. The same situation applied when material from media of whole cells was examined (data not shown). Thus, contrary to the case with β-lactamase, the 46,000 MW protein was present only in very small amounts in the periplasmic space if at all.

Immunological detection of growth hormone synthesised by bacteria

Escherichia coli colonies containing pEx-RGH were screened for their content of RGH using a modification[16] of a solid-phase

immunological method[17] (Fig. 4). Briefly, colonies were lysed *in situ* with $CHCl_3$ and any RGH sequences present in the cells were bound to a polyvinyl disk coated with IgG purified from antiserum to RGH raised in monkeys (see ref. 16 for details on preparation of antiserum). Bound antigen was then specifically labelled with highly purified ^{125}I-anti-RGH IgG (see legend to Fig. 5) and colonies producing RGH were visualised by autoradiography. As shown in Fig. 4, colonies containing the expression plasmid were labelled with the ^{125}I-anti-RGH-IgG probe, whereas those containing pBR322 failed to do so.

Several controls demonstrated the immunological specificity of the assay for RGH and strengthened the conclusion that the clones are producing pituitary GH sequences (Fig. 5). First, colonies of the K12 strain of *E. coli* devoid of plasmid were not labelled with the ^{125}I-anti-RGH-IgG probe. Second, labelling of lysates of colonies containing pEx-RGH was prevented when a large excess (250-fold) of nonradioactive GH was incubated with radioactive probe. Third, labelling was not observed when normal monkey serum or calf serum (data not shown) were used in place of anti-RGH antiserum. Finally, ^{125}I-labelled normal monkey serum failed to label any colony from the strains described above (not shown).

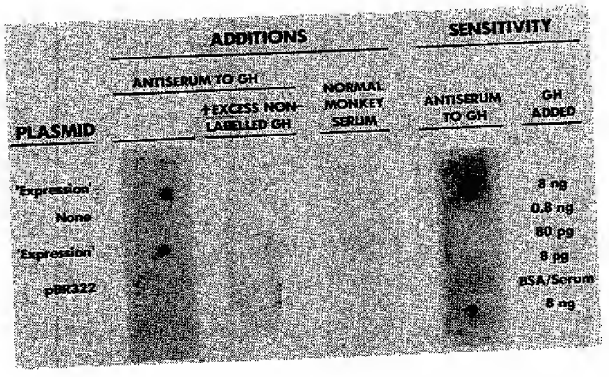


Fig. 5 Immunological detection of GH sequences synthesised by bacteria: specificity of the assay. In 'antiserum to GH', the colonies were treated as described in the legend to Fig. 4. 'None' refers to colonies of *E. coli* devoid of plasmid. In other experiments ('+ excess non-labelled GH'), a large excess (250-fold, 5 μg) of highly purified RGH was preincubated with ^{125}I-anti-RGH IgG. The autoradiograph shown in the lane 'normal monkey serum' results from using polyvinyl strips coated with normal monkey serum (in place of RGH antiserum). A slight reaction occurred with cells containing pEx-RGH, as antigen (RGH) sticks to plastic even in the absence of anti-RGH IgG. The sensitivity of the assay was measured by applying varying amounts of pure RGH (8 ng, 0.8 ng, 80 pg, 8 pg and 8 ng) to an agar plate and screening as described above.

The sensitivity of the screening procedure is indicated in Fig. 5. Labelling was observed with 8 ng, but not with 0.8 ng, of authentic RGH added to the plate. Thus, each colony seems to contain several nanograms of GH sequences, although further experiments are needed to obtain better quantitation.

As indicated in Table 1, immune reactivity was also detected by direct radioimmunoassay in extracts of spheroplasts and in material from the periplasmic space. No such activity was observed in similar isolates from bacteria containing pBR322.

Discussion

In the present studies, bacteria have been genetically programmed to synthesise mammalian pituitary growth hormone sequences. Further analysis is required to assess the fidelity of such expression, the stability of the protein products, and their biological activity. The absolute quantity of GH synthesised by the bacteria cannot be determined accurately from the current studies (an estimate of 24,000 molecules per cell is obtained from the solid-phase immunoassay). However, judging from the prominence of the band in Fig. 3, it seems that sequences of RGH can be a major plasmid product. The quantity of GH present is less than that of β-lactamase; however, the study shown in Fig. 3 may result in an underestimate of the amount of GH synthesised, because in more recent studies (not shown), material of lower molecular weight (in addition to the larger molecular weight material) is specifically immunoprecipitated by antiserum to RGH but not by antiserum to bovine serum albumin. This material could consist of degradation products or the products of premature termination.

The *amp*^{r} gene was used for expressing RGH coding sequences with the hope that the mammalian hormone would be secreted by the bacteria. Such secretion might facilitate the isolation of the protein product and also protect the protein from degradation by proteases. The absence of GH sequences in the periplasmic space is surprising in view of the fact that Villa-Komaroff *et al.*[20] have recently constructed a gene suitable for the expression of rat insulin through linkage with the *amp*^{r} gene and found that the protein product was secreted[20]. It is possible that the hydrophobic region corresponding to the pre-part of RGH in the middle of the hybrid protein prevents the kind of conformation required for transport through membranes. Similarly, the overall structure of the fusion polypeptide might prevent correct processing of the precursor sequence. More detailed studies should help determine which factors, other than pre-sequence of the signal peptide protein of the hormone, are necessary for secretion.

Our findings and those of Villa-Komaroff *et al.*[20] indicate that coding sequences from higher organisms can be expressed in bacteria. Thus, it should be possible to produce biologically important peptides with the use of recombinant DNA techniques and naturally occurring structural gene sequences. Finally, the capability of expression of genes that are designed specifically or synthesised[21] should make possible the production of new hormone analogues for examining the structural requirements for agonist or antagonist activity or for therapy.

H.M.G. and J.D.B. are Investigators of the Howard Hughes Medical Institute. We thank Edmund Tisher for help with these studies. This work was supported in part by NIH grants AM-19997, CA14026 and CA23046, NSF grant PCM 77-16453, and a grant from Lilly.

Received 20 October; accepted 13 November 1978.

1. Ullrich, A. *et al. Science* **196**, 1313–1319 (1977).
2. Seeburg, P. H., Shine, J., Martial, J. A., Baxter, J. D. & Goodman, H. M. *Nature* **270**, 486–494 (1977).
3. Shine, J., Seeburg, P. H., Martial, J. A., Baxter, J. D. & Goodman, H. M. *Nature* **270**, 494–499 (1977).
4. Goodman, H. G., Grumbach, M. M. & Kaplan, S. L. *New Eng. J. Med.* **278**, 57–68 (1968).
5. Tanner, J. M., Whitehouse, R. H., Hughes, P. R. C. & Vince, F. P. *Archs Dis. Child.* **46**, 745–782 (1971).
6. Winawa, S. J., Sherlock, P., Lipkin, M., Sonenberg, M. & Vanamer, P. *Advances in Human Growth Hormone Research*, 862–874, DHEW Publ. No. (NIH) 74–612 (1973).
7. Wallis, M. & Davies, R. V. in *Growth Hormone and Related Polypeptides* (eds Pecile, A. & Muller, E. E.) 1–14 (Elsevier, New York, 1976).
8. Knobil, E. & Greer, R. D. *Recent Prog. Horm. Res.* **15**, 1–12 (1959).
9. Bolivar, F. Rodriguez, Betlach, M. C. & Boyer, H. W. *Gene* **2**, 75–93 (1977).
10. Sutcliffe, J. G. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3737–3741 (1978).
11. Rodriguez, R. L., Bolivar, F., Goodman, H. M., Boyer, H. W. & Betlach, M. in *ICN/UCLA Symp. Molec. Mech. Control Gene Expression* (eds Nierlich, D. P., Rutter, W. J. & Fox, C. F.) 471–477 (Academic, New York, 1976).
12. Roozen, K. J., Fenwick, R. G. Jr & Curtis, R., III *J. Bact.* **107**, 21–33 (1971).
13. Adler, H. I., Fischer, W. D., Cohen, A. & Hardigrer, A. A. *Proc. nam. Acad. Sci. U.S.A.* **57**, 321–326 (1966).
14. Leammli, U. K. *Nature* **227**, 680–685 (1970).
15. Neu, H. C. & Hepel, L. A. *J. biol. Chem.* **240**, 3685–3692 (1965).
16. Ivarie, R. D. & Morris, J. A. (in preparation).
17. Broome, S. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2746–2749 (1978).
18. Miller, J. H. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, New York, 1976).
19. Papavasiliou, S. S., Martial, J. A., Latham, K. R. & Baxter, J. D. *J. clin. Invest.* **60**, 1230–1239 (1977).
20. Villa-Komaroff, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3727–3731 (1978).
21. Itekura, K. *et al. Science* **198**, 1056–1063 (1977).

lette

Discovery identificati

THE discovery o
on SAS 3 at a
$b^{II} \sim -9°$ is repo
revealed a likel
magnitude star
therefore, likely
From 10 to
collimator X-ra
X-ray emission
Data were accur
correlation map
which we desig
$\delta = -63°04'27''$
(90% confidenc
ESO (B) surve
average flux de
2.3 ± 0.3 μJy a
keV^{-1}).
Optical obse
made with the
Observatory (E
possible strong
prime-focus pla
of the sky cen
remarkably bri
survey print th
17.
Spectrogram
with the Carne
Cassegrain spe
1978, 03.33 U
and 20 min, re

Fig. 1 Erro
print.

0028–0836/78/027

TAB P

Proc. Natl. Acad. Sci. USA
Vol. 76, No. 1, pp. 106–110, January 1979
Biochemistry

Expression in *Escherichia coli* of chemically synthesized genes for human insulin

(plasmid construction/*lac* operon/fused proteins/radioimmunoassay/peptide purification)

DAVID V. GOEDDEL*†, DENNIS G. KLEID*, FRANCISCO BOLIVAR*, HERBERT L. HEYNEKER*, DANIEL G. YANSURA*, ROBERTO CREA*‡, TADAAKI HIROSE‡, ADAM KRASZEWSKI‡, KEIICHI ITAKURA‡, AND ARTHUR D. RIGGS†‡

*Division of Molecular Biology, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080; and ‡Division of Biology, City of Hope National Medical Center, Duarte, California 91010

Communicated by Ernest Beutler, October 3, 1978

ABSTRACT **Synthetic genes for human insulin A and B chains were cloned separately in plasmid pBR322. The cloned synthetic genes were then fused to an *Escherichia coli* β-galactosidase gene to provide efficient transcription and translation and a stable precursor protein. The insulin peptides were cleaved from β-galactosidase, detected by radioimmunoassay, and purified. Complete purification of the A chain and partial purification of the B chain were achieved. These products were mixed, reduced, and reoxidized. The presence of insulin was detected by radioimmunoassay.**

Recently improved methods of DNA chemical synthesis, combined with recombinant DNA technology, permit the design and relatively rapid synthesis of modest-sized genes that can be incorporated into prokaryotic cells for gene expression. The feasibility of this general approach was first demonstrated by the synthesis, and expression in *Escherichia coli*, of a gene for the mammalian peptide somatostatin (1).

Following the precursor protein approach used for somatostatin (1), the experimental design for this work was such that the insulin peptide chains would be made *in vivo* as short tails joined by a methionine to the end of β-galactosidase. After synthesis, the insulin chains, which contain no methionine, can be cleaved off efficiently by treatment with cyanogen bromide. We deliberately chose to construct two separate bacterial strains, one for each of the two peptide chains of insulin: the 21-amino-acid A chain and the 30-amino-acid B chain. In native insulin, the two chains are held together by two disulfide bonds, and methods have been available for years for joining the chains correctly, *in vitro*, by air oxidation (2). The efficiency of correct joining has been variable and often low. However, by using *S*-sulfonated derivatives and an excess of A chain, 50–80% correct joining has been obtained (3).

The synthetic plan and chemical synthesis of the DNA fragments coding for the A and B chains of human insulin were described in a previous paper (4) and were summarized in Fig. 1 of that paper. In this communication, we describe the assembly and cloning of the genes for the A and B chains, their insertion into the carboxy terminus of the *E. coli* β-galactosidase structural gene, the expression and purification of the separate A and B chains, and their joining to form native human insulin.

MATERIALS AND METHODS

Bacterial Strains. *E. coli* K-12 strain 294 (*endA*, *thi*$^-$, *hsr*$^-$, *hsm*$_k$$^+$) (5) was provided by K. Backman. *E. coli* K-12 strain D1210, a *lac*$^+$ (*i*^{Q}*o*$^+$*z*$^+$*y*$^+$) derivative of HB101, was constructed by J. Betz and J. Sadler and obtained from J. Sadler.

Enzymes and DNA Preparations. T4 DNA ligase and T4 polynucleotide kinase were purified as described (6). Restriction endonuclease *Eco*RI was purified by the procedure of Greene *et al.* (7); *Hin*dIII was purified by a method developed by D. Goeddel (unpublished). Restriction endonuclease *Bam*HI was purchased from Bethesda Research (Rockville, MD); *E. coli* alkaline phosphatase was purchased from Worthington.

Plasmids, including pBR322 (8), were isolated by a published procedure (9) with some modifications. The chemical synthesis of the deoxyoligonucleotides (figure 1 of ref. 4) has been described (4). λ*plac5* DNA was isolated as described (10).

The following reaction buffers were used: kinase buffer, 60 mM Tris·HCl, pH 8/15 mM 2-mercaptoethanol/10 mM $MgCl_2$; ligase buffer, 20 mM Tris·HCl, pH 7.5/10 mM dithiothreitol/10 mM $MgCl_2$; *Bam*HI buffer, 20 mM Tris·HCl, pH 7.5/7 mM $MgCl_2$/2 mM 2-mercaptoethanol; *Eco*RI–*Hin*dIII buffer, *Bam*HI buffer containing 50 mM NaCl; and phosphatase buffer, 50 mM Tris·HCl, pH 8/10 mM $MgCl_2$.

Assembly of Insulin Genes. The assembly of the right (BB) half of the B-chain gene (see figure 1 of ref. 4) will be described in detail. Oligonucleotides B_2–B_9 were phosphorylated individually. Fifty microcuries of [γ-^{32}P]ATP (≈2000 Ci/mmol, New England Nuclear) was evaporated to dryness in a 1.5-ml polypropylene tube, then incubated with the oligonucleotide (10 μg) and 8 units of T4 polynucleotide kinase in 60 μl of kinase buffer. After 20 min at 37°C, 10 nmol of ATP and 10 units of T4 kinase were added and the reaction was continued for an additional hour. The kinase was inactivated by heating at 90°C for 5 min.

Phosphorylated fragments B2, B3, B6, and B7 (2.5 μg each) were combined with 2.5 μg of 5′-OH fragment B1 and dialyzed for 2 hr against 1 liter of ligase buffer. ATP was added to a concentration of 0.2 mM, the reaction mixture (60 μl) was cooled to 12°C, and T4 ligase (50 units) was added. A separate ligation reaction involving phosphorylated fragments B4, B5, B8, and B9 and unphosphorylated B10 was performed identically. After 12 hr at 12°C, the two ligation reaction mixtures were combined, additional T4 ligase (40 units) was added, and the mixture was incubated at 12°C for 4 hr. The mixture was extracted with phenol/chloroform and precipitated with ethanol, and the DNA fragments were purified by electrophoresis on a 10% acrylamide gel (11). The most slowly migrating band was sliced from the gel and the DNA was extracted (11).

A similar procedure, with the following exceptions, was used to assemble the left (BH) half of the B-chain gene. All eight

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: BB, left half of insulin B gene; BH, right half of insulin B gene; A(SSO_3^-), *S*-sulfonated derivatives of the insulin A-chain peptide; B(SSO_3^-), *S*-sulfonated derivatives of the insulin B-chain peptide.

† To whom reprint requests should be addressed.

oligonucleotides (H1–H8 in figure 1 of ref. 4, 30 μg each) were phosphorylated. Therefore, after complete ligation and before purification by gel electrophoresis, the reaction mixture was treated with 400 units of *Eco*RI and 400 units of *Hin*dIII for 2 hr at 37°C. The BH band migrating at 46 base pairs was eluted from a 10% acrylamide gel.

The procedure used to construct the A-chain gene was also similar to that described for the BB fragment. The only major difference was that, after assembly, the 5′ ends of the complete A gene were phosphorylated.

Construction and Characterization of *lac*–Insulin Hybrid Plasmids. The BB fragment was cloned as follows: 1 μg of pBR322 was treated with 5 units of *Bam*HI in *Bam*HI buffer for 1 hr at 37°C. After addition of NaCl to 50 mM, *Hin*dIII (5 units) was added and the reaction was continued for 1 hr. The enzymes were inactivated by heating at 70°C for 10 min. The prepared pBR322 was ligated to the BB fragment for 3 hr at 12°C in 25 μl of ligase buffer (containing 0.16 mM ATP) by using 20 units of T4 ligase. Half of the resulting DNA was used to transform *E. coli* 294 by a published procedure (12). The BH fragment and the A-chain gene were cloned similarly, with the appropriate restriction endonucleases to cut pBR322.

Construction of the plasmids for expression of the synthetic insulin genes is described in the legend of Fig. 1. The separate chains in insulin are biologically inactive (2) and were synthesized attached to large precursor proteins. Therefore, the containment level of P2-EK1, recommended by the National Institutes of Health guideline, was used.

DNA Sequences. The method of Maxam and Gilbert (11) was used to determine DNA sequences. Sequence data are not included, but will be provided upon request.

Preparation of Insulin Reagents. Porcine and bovine insulin were purchased from Sigma. The *S*-sulfonated derivatives (SSO_3^-) of their A and B chains were prepared and purified as described (13). ^{35}S-Labeled A(SSO_3^-) and B(SSO_3^-) were prepared similarly except that 5 mCi of sodium [^{35}S]sulfite (69 mCi/mol, New England Nuclear) was substituted for unlabeled sodium sulfite. After separation of the chains, the specific activity was 92,000 cpm/μg and 32,000 cpm/μg, respectively, for A and B chains. The radioimmunoassay for the insulin chains is described in the legend of Fig. 3.

Purification of B Chain of Human Insulin. *E. coli* D1210/pIB1 was grown to late logarithmic phase in 7 liters of LB medium (10) containing 20 mg of ampicillin per liter. Isopropyl-β-D-thiogalactoside was added to a final concentration of 2 mM, and the cells were grown for one more doubling. Wet cell paste (24 g) was suspended in 30 ml of BB buffer (10) and cells were lysed by one passage through a French press at 4000 lb/inch^2 (27.6 MPa). The cell debris was pelleted by centrifugation at 15,000 rpm for 30 min. The pellet was dissolved in 40 ml of 6 M guanidinium chloride/1% 2-mercaptoethanol and centrifuged at 30,000 rpm for 1 hr. The supernatant was dialyzed overnight against 20 liters of H_2O. The precipitate, containing about 1 g of protein, was dissolved in 25 ml of 70% formic acid. Cyanogen bromide (1.3 g) was added and the mixture was allowed to react overnight at room temperature. Formic acid and cyanogen bromide was removed by rotary evaporation and the residue was dissolved in 50 ml of 8 M guanidinium chloride. *S*-Sulfonated derivates of the peptide mixture were prepared by adding 1 g of sodium tetrathionate and 2 g of sodium sulfite, adjusting the pH to 9 with NH_4OH, and stirring the mixtures at room temperature for 24 hr. The pH was then adjusted to 5 with acetic acid and the mixture was dialyzed twice against 3 liters of H_2O. The resulting white precipitate (≈0.6 g of protein) was pelleted by centrifuging at 10,000 rpm for 10 min.

Purification of A Chain of Human Insulin. *E. coli* 294/pIA1 was grown to A_{550} of 2 in 5 liters of LB medium containing 20 mg of ampicillin per liter. This strain is constitutive for β-galactosidase and so was not induced. The 15 g (wet weight) of cells obtained were processed by the same procedure used for the B chain up through the preparation of the *S*-sulfonated derivatives. After the pH was adjusted to 5 and the mixture was dialyzed against H_2O to an ionic strength of about

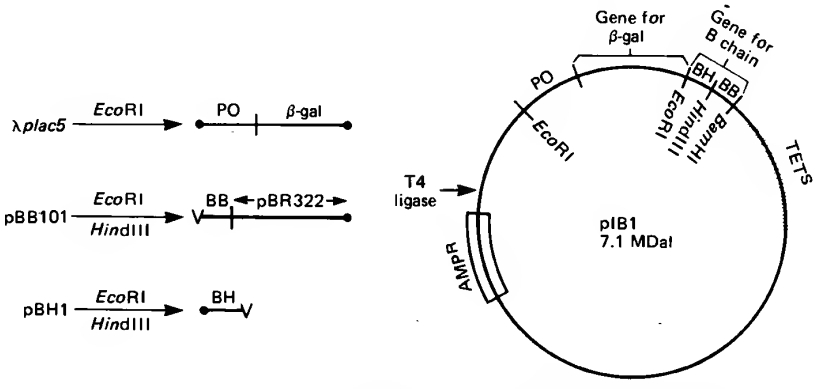


FIG. 1. Construction of *lac*–insulin plasmids. pBB101 (2 μg) (pBR322 containing the BB sequence) was treated with *Eco*RI and *Hin*dIII (20 units each), and the large fragment was purified on a 10% acrylamide gel. pBH1 (8 μg) (pBR322 containing BH sequence) also was treated with *Eco*RI and *Hin*dIII, and the small fragment was purified on a 10% acrylamide gel. These two fragments were ligated to 2 μg of *Eco*RI-digested λ*plac5* in 30 μl of ligase buffer with 20 units of ligase. This mixture was used to transform *E. coli* 294. The configuration of restriction site ends (V represents *Hin*dIII; ● represents *Eco*RI) was such that only correct joining of the two halves of the B gene would lead to viable plasmids. To screen for the presence of the *lac* fragment, we plated the transformed culture on glucose minimal plates (10) containing 40 μg of 5-bromo-4-chloro-3-indoyl-β-galactoside (X-gal) and 20 μg of ampicillin per ml. Plasmids were prepared from β-galactosidase constitutive (blue) colonies. Because the λ*plac5 lac* operon fragment contains an asymmetrical *Hin*dIII site (14), the orientation of that fragment in the resulting plasmids can be determined. Plasmid samples of 1 μg were digested with *Hin*dIII and sized on 0.7% agarose gels. Plasmids (15-μg samples) having the desired orientation of the *lac* fragment were then treated with *Eco*RI, *Hin*dIII, and *Bam*HI, and sized on a 10% acrylamide gel to verify the presence of both the BH and BB fragments. The diagram of pIB1 (7.1 megadaltons) is not drawn to scale. To construct the *lac*–insulin A plasmid (pIA1, not shown), we ligated 1 μg of *Eco*RI-treated pA11 (pBR322 containing the A gene) and 3 μg of *Eco*RI-treated λ*plac5* for 4 hr at 4°C. Transformants of *E. coli* 294 were selected for resistance to ampicillin on X-gal plates. Orientation of the *lac* fragment was determined by digesting plasmids purified from the blue colonies with *Hin*dIII and *Bam*HI.

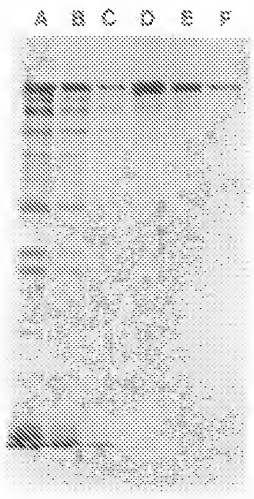


FIG. 2. Sodium dodecyl sulfate/polyacrylamide gel electrophoresis of extracts of strain 294/pIA1. Samples were heated in O'Farrell sample buffer and electrophoresed in a sodium dodecyl sulfate/10% gel as described (15). Lanes A, B, and C: total cells; 20, 10, and 5 μl, respectively. Lanes D, E, and F: insoluble cell debris; 20, 10, and 5 μl, respectively.

0.01 M, the mixture was centrifuged and the supernatant was used for further purification (see *Results and Discussion*).

RESULTS AND DISCUSSION

Assembly and Cloning of B-Chain Gene and A-Chain Gene. The gene for the B chain of insulin was designed to have an *Eco*RI restriction site on the left end, a *Hin*dIII site in the middle, and a *Bam*HI site at the right end. This was done so that both halves, the left *Eco*RI–*Hin*dIII half (BH) and the right *Hin*dIII–*Bam*HI half (BB), could be separately cloned in the convenient cloning vehicle pBR322 (8) and, after their sequences had been verified, joined to give the complete B gene (Fig. 1). The BB half was assembled by ligation from 10 oligodeoxyribonucleotides, labeled B1–B10 in figure 1 of ref. 4, made by phosphotriester chemical synthesis. B1 and B10 were not phosphorylated, thereby eliminating unwanted polymerization of these fragments through their cohesive ends (*Hin*dIII and *Bam*HI). After purification by preparative acrylamide gel electrophoresis and elution of the largest DNA band, the BB fragment was inserted into plasmid pBR322 that had been cleaved with *Hin*dIII and *Bam*HI. About 50% of the ampicillin-resistant colonies derived from the DNA were sensitive to tetracyline, indicating that a nonplasmid *Hin*dIII–*Bam*HI fragment had been inserted. The sequences of the small *Hin*dIII–*Bam*HI fragments from four of these colonies (pBB101 to pBB104) were determined (11) and were correct as designed.

The BH fragment was prepared in a similar manner and inserted into pBR322 that had been cleaved with *Eco*RI and *Hin*dIII restriction endonucleases. Plasmids from three ampicillin-resistant, tetracycline-sensitive transformants (pBH1 to pBH3) were analyzed. The small *Eco*RI–*Hin*dIII fragments had the expected nucleotide sequence.

The A-chain gene was assembled in three parts. The left four, middle four, and right four oligonucleotides (see figure 1 of ref. 4) were ligated separately, then mixed and ligated (oligonucleotides A1 and A12 were unphosphorylated). The assembled A-chain gene was phosphorylated, purified by gel electrophoresis, and cloned in pBR322 at the *Eco*RI–*Bam*HI sites. The

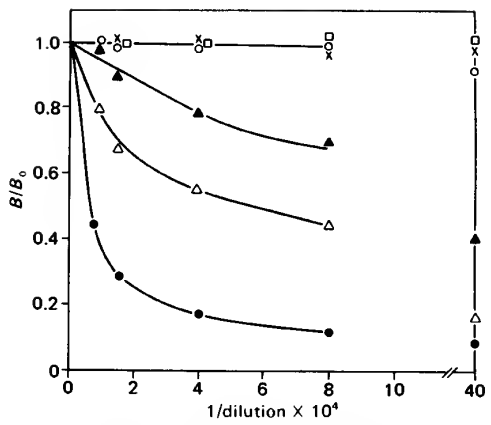


FIG. 3. Reconstitution radioimmunoassay for insulin chains. The *S*-sulfonated A sample was mixed with the *S*-sulfonated B sample in a 1.5-ml conical polypropylene tube and dried under reduced pressure. The dried proteins were suspended in 25 μl of 10 mM sodium acetate (pH 4.5). Two microliters of 10% (vol/vol) 2-mercaptoethanol was added and the mixture was heated (≈95°C) for 10 min. The mercaptoethanol was removed by ethyl acetate extraction. Two microliters of 0.1 M glycine buffer (pH 10.6) was added, and the pH was adjusted, if necessary, to 9.6–10.6. The open tube was placed in a dessicator over moist hyamine hydroxide at room temperature for ≥6 hr. A diluted aliquot of the reaction mixture was assayed for insulin radioimmune activity by use of a commercially available radioimmunoassay kit (Phadebus Insulin Test, Pharmacia). Dilution and assay were done according to the instructions supplied. The ordinate B/B_0 is the cpm in the pellet of the experimental sample divided by the cpm in the pellet obtained with buffer only. ●, 40 μg of porcine $A(SSO_3^-)$ in 10 mM NH_4HCO_3 (pH 9) was mixed with 10 μg of bovine $B(SSO_3^-)$ in the same buffer; ○, porcine $A(SSO_3^-)$ only; □, bovine $B(SSO_3^-)$ only; ▲, 100 μg of porcine $A(SSO_3^-)$ and 93 μg of *E. coli* B-chain fraction F-10 (cleaved by CNBr and *S*-sulfonated, insoluble at pH 5); ×, fraction F-10 only; △, 100 μg of porcine $A(SSO_3^-)$, 93 μg of fraction F-10, and 3 μg of bovine $B(SSO_3^-)$.

*Eco*RI–*Bam*HI fragments from two ampicillin-resistant, tetracycline-sensitive clones (pA10 and pA11) contained the desired A-gene sequence.

Construction of Plasmids for Expression of A and B Insulin Genes. Fig. 1 illustrates the construction of the *lac*-insulin B plasmid (pIB1). Plasmids pBH1 and pBB101 were digested with *Eco*RI and *Hin*dIII endonucleases. The small BH fragment of pBH1 and the large fragment of pBB101 (containing the BB fragment and most of pBR322) were purified by gel electrophoresis, mixed, and ligated in the presence of *Eco*RI-cleaved λ*plac5*. The 4.4-megadalton *Eco*RI fragment of λ*plac5* contains the *lac* control region and the majority of the β-galactosidase structural gene (1, 14). The configuration of the restriction sites ensures correct joining of BH to BB. The *lac Eco*RI fragment can be inserted in two orientations; thus, only half of the clones obtained after transformation should have the desired orientation. The orientation of 10 ampicillin-resistant, β-galactosidase-constitutive clones were checked by restriction analysis (see legend of Fig. 1). Five of these colonies contained the entire B-gene sequence and the correct reading frame from the β-galactosidase gene into the B-chain gene. One, pIB1, was chosen for subsequent experiments.

In a similar experiment, the 4.4-megadalton *lac* fragment from λ*plac5* was introduced into the pA11 plasmid at the *Eco*RI site to give pIA1. pIA1 is identical to pIB1 except that the A-gene fragment is substituted for the B-gene fragment. DNA sequence analysis showed that the correct A- and B-chain gene sequences were retained in pIA1 and pIB1, respectively.

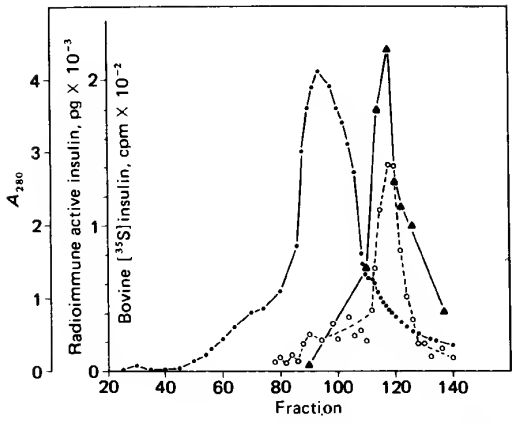


FIG. 4. DEAE-cellulose chromatography of an extract of strain D1210/pIB1. The fraction that was insoluble at pH 5 (580 mg of protein) was dissolved in 20 ml of 10 mM NH_4HCO_3 and adjusted to pH 9. ^{35}S-Labeled bovine B(SSO_3^-) (135,000 cpm; 4.4 μg) was added and the sample was applied to a 2 × 60 cm column of Whatman DE52. Elution was with a 1-liter gradient of 0.01–2.0 M NH_4HCO_3 (pH 9). Fractions of 4 ml were collected. ●, A_{280}; ○, 100-μl aliquots were used to measure radioactivity of ^{35}S-labeled bovine B(SSO_3^-); ▲, 100-μl aliquots were assayed for B-chain radioimmune activity by being mixed with 100 μg of porcine A(SSO_3^-) and using the reconstitution assay (Fig. 3).

Expression. The strains that contain the insulin genes correctly attached to β-galactosidase (D1210/pIB1 and 294/pIA1) both produce large quantities of a protein the size of β-galactosidase (Fig. 2). Approximately 20% of the total cellular protein was this β-galactosidase–insulin A or B chain hybrid. The hybrid proteins are insoluble and were found in the first low-speed pellet in which they constitute ≈50% of the protein (Fig. 2).

To detect the expression of the insulin A and B chains, we used a radioimmunoassay based on the reconstitution of complete insulin from the separate chains. The insulin reconstitution procedure of Katsoyannis *et al.* (3), adapted to a 27-μl assay volume, provided a very suitable assay. Easily detectable insulin radioimmune activity was obtained after S-sulfonated derivatives of the insulin chains were mixed and reconstituted by the procedure described in the legend of Fig. 3. The separate S-sulfonated chains of insulin do not react significantly, after reduction and oxidation, with the anti-insulin antibody used. Our reconstitution assay, though not extremely sensitive (limits of detection about 1 μg), was specific and suitable for following insulin chain radioimmune activity during purification.

To use the reconstitution assay, we partially purified the β-galactosidase–A or B chain hybrid protein, cleaved it with cyanogen bromide, formed *S*-sulfonated derivatives, and partially purified the peptides as described in *Materials and Methods*. This procedure was based on our earlier experience with the purification of somatostatin from *E. coli* (unpublished data) and the known properties of the insulin chains.

Table 1. Amino acid composition of *E. coli* insulin A chain

Amino acid	Residues per peptide		
	E. coli A(SSO_3^-)	Porcine A(SSO_3^-)	Predicted
His	0.08	0.08	0
Lys	0.00	0.00	0
Trp	0.00	0.00	0
Arg	0.00	0.00	0
Phe	0.00	0.00	0
Asx	2.38	2.50	2
Thr	0.24	0.28	1
Ser	0.14	0.23	2
H-Ser	0.02	0.00	0
Glx	3.97	4.58	4
Pro	0.00	0.09	0
Gly	1.40	1.48	1
Ala	0.20	0.11	0
Cys	0.55	0.00	0
Val	1.15	1.06	1
Met	0.62	0.43	0
Ile	1.99	1.48	2
Leu	2.33	2.35	2
Tyr	1.89	2.30	2

Approximately 25 μg of porcine A(SSO_3^-) (which is identical in sequence to human A) and 25 μg of *E. coli* A(SSO_3^-) purified twice by high-performance liquid chromatography were hydrolyzed and analyzed in parallel. The SSO_3^- derivatives of cysteine were destroyed during hydrolysis and do not register as amino acids with the program used. Serine and threonine were also partially destroyed.

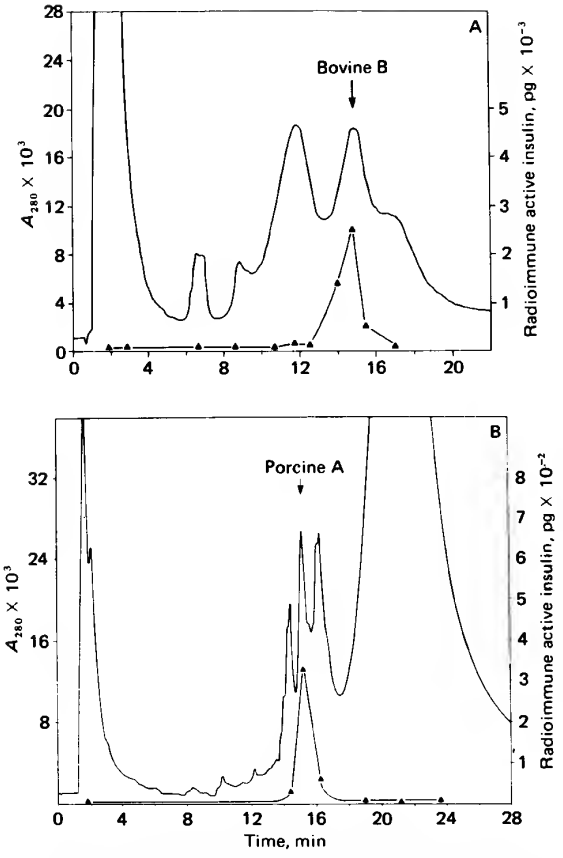


FIG. 5. Reversed-phase high-performance liquid chromatography. (*A*) A sample (500 μg), purified by DEAE-cellulose chromatography (B chain, Fig. 4), was subjected to high-performance liquid chromatography at room temperature on an SP 3500 liquid chromatograph (Spectra-Physics) equipped with a LiChrosorb RP-8 0.3 × 25 cm column (Merck EM). The eluting buffer was 50 mM NH_4OAc with an acetonitrile gradient of 24–60%. Fractions were collected and dried, and samples were assayed for B-chain radioimmune activity by our reconstitution assay. (*B*) An A-chain sample (500 μg total protein), partially purified by aminoethyl-cellulose chromatography, was subjected to high-performance liquid chromatography with an acetonitrile gradient of 15–60%. Fractions were collected and samples were assayed for A-chain radioimmune activity. Solid line, A_{280}; ▲, radioimmune activity.

Table 2. Reconstitution of radioimmune human insulin

"A" sample	"B" sample	Radioimmune active insulin, ng
E. coli 58-HPLC*	—	<0.5
—	*E. coli* DE117†	<0.5
Porcine A‡	*E. coli* DE117	74
E. coli 58-HPLC	Bovine B§	45
E. coli 58-HPLC	*E. coli* DE117	20

Our standard reconstitution assay procedure was used (Fig. 3). The results are given as ng of radioimmune active insulin produced per 20 μl of the reaction mixture. HPLC, high-performance liquid chromatography.

* Five hundred microliters of fraction 58 from an aminoethyl-cellulose column was chromatographed on an RP-8 column and the "A" peak was collected. As estimated from the peak height, the sample contained approximately 25 μg of protein.

† Ten microliters of DEAE-cellulose fraction 117 (Fig. 4), concentrated to 1.6 mg of total protein per ml, was used as the "B" sample.

‡ *S*-Sulfonated porcine A (70 μg).

§ *S*-Sulfonated bovine B (10 μg).

Insulin B-chain radioimmune activity was detected first among the *S*-sulfonated cyanogen bromide peptides insoluble at pH 5 (fraction F-10, Fig. 3). The activity was enriched further by chromatography on DEAE-cellulose (Fig. 4). The B-chain radioimmune activity coeluted with *S*-[^{35}S]sulfonated bovine B chain.

A portion of the material purified by DEAE-cellulose chromatography was analyzed by high-performance liquid chromatography on a reversed-phase RP-8 column (Fig. 5*A*). This column separates primarily on the basis of hydrophobic interactions. The insulin B-chain radioimmune activity eluted at a position very close to that of bovine B chain. Good purification was obtained by high-performance liquid chromatography, but the breadth of the peak indicated that the chromatographic fraction was not pure.

Another sample (1 mg total protein) of the material purified by DEAE-cellulose chromatography was subjected to gel filtration on Sephadex G-75 in 50% acetic acid, a system that completely resolves A chain from B chain. The B-chain radioimmune activity eluted at the same position as *S*-sulfonated bovine B chain, indicating similar sizes (data not shown).

Insulin A-chain radioimmune activity was detected first in the total mixture of cyanogen bromide peptide fragments obtained from the partially purified β-galactosidase–A chain hybrid. The activity was enriched by pH 5 precipitation and aminoethyl-cellulose chromatography and purified on a microgram scale by high-performance liquid chromatography (Fig. 5*B*). The insulin A-chain radioimmune activity eluted from the column at a position indistinguishable from that of porcine *S*-sulfonated A chain. Porcine A chain is identical to human A chain (2).

When an excess of porcine A(SSO_3^-) (40 μg) was mixed, reduced, and oxidized with bovine B(SSO_3^-) (10 μg), we usually obtained 10–15% correct joining to yield radioimmune active insulin. Reconstitution in impure mixtures was lower, as expected. Because of this strong and variable competitive inhibition by other peptides, the amount of insulin chains in the extracts can best be estimated by adding to the extract a known amount of the chain to be assayed. This type of experiment (illustrated in Fig. 3) indicates that the yield of insulin chains is high (approximately 10 mg from 24 g wet weight of cells) and consistent with the amount of insoluble β-galactosidase protein obtained (at least 10^5 molecules per cell). This estimated yield is 10 times higher than that reported for somatostatin (1).

The evidence that we have obtained correct expression from chemically synthesized genes for human insulin can be summarized as follows. (*i*) Radioimmune activity has been detected for both chains. (*ii*) The DNA sequences obtained after cloning and plasmid construction have been directly verified to be correct as designed. Because radioimmune activity is obtained, translation must be in phase. Therefore, the genetic code dictates that peptides with the sequences of human insulin are being produced. (*iii*) The *E. coli* products, after cyanogen bromide cleavage, behave as insulin chains in three different chromatographic systems that separate on different principles (gel filtration, ion exchange, and reversed-phase high-performance liquid chromatography). (*iv*) The A chain produced by *E. coli* has been purified on a small scale by high-performance liquid chromatography and has the correct amino acid composition (Table 1).

Table 2 illustrates that insulin radioimmune activity can be obtained entirely from *E. coli* products. Easily detectable radioimmune insulin activity is produced when purified *E. coli* A chain is mixed and reconstituted with partially purified (≈10% pure) *E. coli* B chain.

We gratefully acknowledge the excellent technical assistance of Louise Shively, Rochelle Sailor, Frances Fields, and Mark Backer. We give special thanks to Herbert W. Boyer for his encouragement and scientific consultation and to Robert A. Swanson for making this work possible. Work at the City of Hope was supported by contracts from Genentech, Inc.

1. Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F. & Boyer, H. W. (1977) *Science* **198**, 1056–1063.
2. Humbel, R. E., Bosshard, H. R. & Zahn, H. (1972) in *Handbook of Physiology*, eds. Steiner, D. F. & Freinkel, N. (Williams & Wilkins, Baltimore, MD), Vol. 1, pp. 111–132.
3. Katsoyannis, P. G., Trakatellis, A. C., Johnson, S., Zalut, C. & Schwartz, G. (1967) *Biochemistry* **6**, 2642–2655.
4. Crea, R., Hirose, T., Kraszewski, A. & Itakura, K. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5765–5769.
5. Backman, K., Ptashne, M. & Gilbert, W. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 4174–4178.
6. Panet, A., van de Sande, J. H., Loewen, P. C., Khorana, H. G., Raae, A. J., Lillenhaug, J. R. & Kleppe, K. (1973) *Biochemistry* **12**, 5045–5049.
7. Greene, P. J., Betlach, M., Bolivar, F., Heyneker, H. L., Tait, R. & Boyer, H. W. (1978) *Nucleic Acids Res.* **12**, 2373–2380.
8. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H. & Falkow, S. (1977) *Gene* **2**, 95–119.
9. Clewell, D. B. (1972) *J. Bacteriol.* **110**, 667–676.
10. Miller, J. H. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
11. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
12. Hershfield, V., Boyer, H. W., Yanofsky, C., Lovett, M. A. & Helinski, D. R. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 3455–3459.
13. Katsoyannis, P. G., Tometsko, A., Zalut, C., Johnson, S. & Trakatellis, A. C. (1967) *Biochemistry* **6**, 2635–2642.
14. Polisky, B., Bishop, R. J. & Gelfand, D. H. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 3900–3904.
15. O'Farrell, P. H. (1975) *J. Biol. Chem.* **250**, 4007–4021.

TAB Q

Gene, 5 (1979) 291—303

CLONING AND EXPRESSION OF THE YEAST GALACTOKINASE GENE IN AN *Escherichia coli* PLASMID

(Galactose gene cluster; *Bgl*II; *Bam*HI; pBR322 vector; mRNA colony hybridization; gene bank; *Saccharomyces cerevisiae*; recombinant DNA)

MARK A. SCHELL and DAVID B. WILSON

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853 (U.S.A.)

(Received November 28th, 1978)
(Accepted February 21st, 1979)

SUMMARY

This report describes the construction and isolation of a plasmid, derived from pBR322, which carries a *Bgl*II restriction fragment of DNA containing the galactokinase gene from *Saccharomyces cerevisiae*. This was accomplished by the following procedure: (1) Purified galactokinase mRNA, labelled with ^{125}I, was hybridized to *Bgl*II digests of yeast DNA employing Southern's filter transfer technique to identify a restriction fragment containing the galactokinase gene. (2) This fragment was partially purified by agarose gel electrophoresis, ligated into the *Bam*HI site of pBR322 and transformed into *Escherichia coli* to generate a clone bank containing the galactokinase gene. (3) This bank was screened by in situ colony hybridization with galactokinase mRNA resulting in the identification of a plasmid carrying this gene. This plasmid DNA hybridized with the galactokinase mRNA to the same extent in the presence or absence of a large excess of unlabelled mRNA from cells that were not induced for galactokinase synthesis, while the same amount of unlabelled galactose-induced mRNA reduced the hybridization by 95%. When this plasmid was introduced into an *E. coli* strain deleted for the galactose operon it caused the synthesis of low levels of yeast galactokinase activity.

INTRODUCTION

We have previously reported the purification of the *gal 1* gene product (galactokinase) from *S. cerevisiae* (Schell and Wilson, 1977). A specific antibody was prepared against this protein and subsequently used to purify the

Abbreviations: BSA, bovine serum albumin; DTT, dithiothreitol; SDS, sodium dodecyl sulphate.

mRNA coding for galactokinase by immunoprecipitation (Schell and Wilson, 1978). As the next step the mRNA was used as a probe to detect a chimeric plasmid containing the yeast galactokinase gene. The successful isolation of this plasmid would demonstrate the utility of this type of sequential isolation (protein → antibody → mRNA → DNA) to obtain probes to study and clone many other eucaryotic genes, whose mRNA's represent less than 2% of the total cellular mRNA.

There are two possible approaches to the cloning of a specific gene fragment utilizing a purified mRNA as a probe for the gene sequence. The first approach, the "shot gun" method, involves the screening of a clone bank containing the entire yeast genome for a clone that would hybridize to galactokinase mRNA. However this approach is complicated by the large number of rDNA clones that are present in a bank of the entire yeast genome as they will hybridize to the small amounts of rRNA which contaminate most mRNA preparations.

Alternatively, the fragment containing the desired gene can be identified by the use of the method of Southern (1973) and partially purified before the generation of the clone bank. This method has several advantages. First the partially purified clone bank is much smaller than the "shot gun" bank and therefore is much easier to generate and screen. Second, hybridization conditions can be established for the identification of the specific gene fragment. Finally, ribosomal DNA can be eliminated from the clone bank by the choice of the proper restriction enzyme to generate the fragment.

METHODS AND PROCEDURES

Preparation of ^{125}I-labelled galactokinase mRNA

Galactokinase mRNA was purified by immunoprecipitation of total yeast polysomes with galactokinase antibody, deproteinization of the RNA in the immunoprecipitated polysomes, and chromatography of this RNA on polyuridylic acid linked Sepharose by the procedure described by Schell and Wilson (1979). In vivo ^{32}P-labelled mRNA prepared by this method did not have a high enough specific activity to be used to probe for the galactokinase gene in restriction enzyme digests of the entire yeast genome. Therefore, the galactokinase mRNA was purified after low level ^{32}P labelling (in vivo; for monitoring purification) and iodinated with ^{125}I by a modification of the method described by Prensky (1976). The reaction was prepared by the addition of the following reagents: 3.0 μl Na ^{125}I (0.4 Ci/ml in 50 mM NaOH; carrier free); 2.0 μl 0.15 N HNO_3 in 0.2 M NaAc (sodium acetate) pH 4.8; 1.5 μl 0.2 mM KI in 0.1 M NaAc pH 4.8; 5.0 μl RNA (0.2—0.6 μg) in 0.1 M NaAc pH 4.8; 2.0 μl 0.1 M NaAc pH 4.8; 5.0 μl 1.4 mM $TlCl_3$ in 0.1 M NaAc pH 4.8.

After the $TlCl_3$ catalyst solution was added, the solution was incubated at 60°C for 25 min; 0.2 ml of TNE (0.1 M Tris·HCl pH 8.0, 0.1 M NaCl, 1 mM EDTA) was added to the reaction and the incubation was continued for an additional 12 min. Then 0.05 mg of *E. coli* rRNA was added to the

solution and the iodinated mRNA was purified away from the unreacted ^{125}I by chromatography on 0.5 ml columns of CF-11 cellulose as described by Prensky (1976). The resultant RNA sample contained $3 \cdot 10^7$ cpm giving it a specific activity of $0.5—1.0 \cdot 10^8$ cpm/μg RNA.

Restriction enzyme analysis of yeast DNA for the galactokinase gene

Yeast DNA was prepared by the method described by Cryer et al. (1975) and further purified by equilibrium sedimentation centrifugation in CsCl. The DNA was digested by placing it in 10 mM Tris·HCl pH 7.5, 10 mM $MgCl_2$, 6 mM DTT, 50 mM NaCl, 0.1 mg/ml BSA to a final DNA concentration of 0.05 to 0.3 mg/ml. Restriction enzyme was added to the solution at a ratio of 1 unit of enzyme to 1 μg of DNA and it was incubated at 37°C for 2—5 h. These samples (5—50 μl) were loaded onto 1% agarose slab gels (0.3 cm × 20 cm × 20 cm) containing 50 mM Tris—Ac pH 8.2, 20 mM NaAc, 20 mM NaCl, 3 mM EDTA. The gels were run at 50 mA for 10 h while circulating the electrophoresis buffer between the upper and lower tanks with a peristaltic pump (Wu et al., 1976).

The restriction fragments separated on these gels were transferred onto nitrocellulose sheets by a modification of the method of Southern (1975). Hybridization of ^{125}I-labelled galactokinase mRNA to these sheets was performed in "Seal-A-Meal" bags in a minimal volume of 5 × SSC + 0.5% SDS at 65°C for 20—30 h (SSC = 0.15 M NaCl, 0.015 M Na-citrate, pH 7.0). Usually 3 ml of this hybridization solution containing 0.3 mg of *E. coli* tRNA and $0.5—2.0 \cdot 10^6$ cpm of ^{125}I-labelled galactokinase mRNA was used for a 10 × 13 cm sheet. After hybridization the nitrocellulose sheet was removed from the bag, washed 3 times with 200 ml of 2 × SSC + 0.5% SDS at 43°C for 30 min each time, and then finally washed with 200 ml of 2 × SSC for 30 min at 43°C three more times. The sheet was air-dried on a glass plate, coated with a solution of 20% PPO in toluene and dried again. The sheets were autoradiographed at −20°C using Kodak XR-5 X-ray film with a Picker intensifying screen for 2—14 days.

Preparation of plasmid DNA

E. coli strains containing the plasmid pBR322 (Rodriguez et al., 1977) were grown at 37°C in M9 media supplemented with 0.2% glucose, 0.3% casamino acids, 0.1 mg/ml ampicillin, and any specific strain requirements to late log phase ($A_{600\,nm} = 0.9$). Chloramphenicol was added to the culture at a final concentration of 0.17 mg/ml and the culture was shaken at 37°C for 12—16 h to amplify the plasmid. Plasmid DNA was prepared from these cells by a modification of the method of Tanaka and Weisblum (1975). The amount of DNA was determined by reading the $A_{260\,nm}$ and by the diphenylamine assay (Burton, 1968).

Extraction of DNA from agarose gels

DNA was extracted from agarose gel sections by the following method. Gel

sections (1.5 ml) were slowly frozen and thawed 3 times and homogenized in a siliconized dounce homogenizer. 3 ml of TNE (20 mM Tris·HCl pH 7.5, 0.4 M NaCl, 5 mM EDTA) were added and the suspension rehomogenized. 3 ml of TNE-saturated phenol were added, the suspension was rehomogenized, allowed to sit at room temperature for 30 min, and centrifuged at 9000 rev./min in an HB-4 rotor. After removal of the aqueous phase the phenol and agarose were extracted 2 more times as described above with 3 ml TNE. The combined aqueous phases were extracted with TNE-saturated phenol and ethanol-precipitated. Approx. 50% of the DNA in the gel sections was recovered by this method.

Ligation and transformation

Ligation of *Bgl*II restriction enzyme fragments into pBR322 plasmid DNA was performed by a modification of the method of Tanaka and Weisblum (1975). The conditions are described as follows. Plasmid pBR322 cut with *Bam*HI restriction enzyme was diluted to 22 μg/ml in the following solution: 30 mM Tris·HCl pH 7.5, 0.05 mM ATP, 12 mM $MgCl_2$, 50 mM NaCl, 10 mM DTT. *Bgl*II restriction enzyme fragments isolated from agarose gels were added at a final concentration of 80 μg/ml; 0.3 units of T4 ligase were added, and the reaction incubated for 12 h at 12°C. The ligation of the yeast DNA fragments from the gel into the plasmid was monitored by agarose gel electrophoresis in a 1% agarose gel containing 90 mM Tris·borate pH 8.3, 3 mM EDTA (Greene et al., 1975).

Competent cells of *E. coli* strain M94 (CGSH/5346) were prepared as described by Morrison (1977). Transformation of these cells with plasmid DNA was performed by a modification of the method of Tanaka and Weisblum (1975). The ligation reaction mix (0.05 ml) was added to 0.06 ml of 10 mM Tris·HCl pH 7.5, 1 mM EDTA, 20 mM NaCl, and 100 mM $CaCl_2$ and 0.2 ml of competent cells were added to this solution. After incubation at 4°C for 30 min the cells and DNA solution were incubated at 42°C for 3 min with shaking. Then 2.7 ml of L-broth + 0.2% glucose + 0.1 mg/ml ampicillin were added to the cell suspension and incubated at 37°C with moderate shaking for 40 min. Aliquots (0.1 ml = 150 transformants) were spread on L-broth plates containing 0.1 mg/ml ampicillin and 0.2% glucose, and the plates incubated at 37°C for 20 h.

The chilled plates were replica plated (Miller, 1972), onto L-broth + 0.2% glucose plates containing 0.02 mg/ml tetracycline, and these plates incubated for 15 h at 37°C. The original master plate (ampicillin) was superimposed onto the tetracycline plate and the colonies that did not grow on the tetracycline plates were picked onto new plates (L-broth + 0.2% glucose + 0.1 mg/ml ampicillin).

Screening of plasmid containing strains for the galactokinase gene

The strains which were found to contain plasmid pBR322 with inserted yeast DNA (ampicillin resistant/tetracycline sensitive) were grown at 37°C

for 15 h on L-broth plates containing 0.1 mg/ml ampicillin. Filters which contained plasmid DNA from each individual strain at the site of each colony were prepared on Whatman 540 paper by a modification of the method of Beckmann et al. (1977). To detect the clones containing the galactokinase gene, the filters were hybridized with ^{125}I-labelled galactokinase mRNA. 20 filters (6 × 8 cm, containing 48 colonies each) were placed in a "Seal-a-Meal" bag and then 15 ml of 50% formamide, 4 × SSC, 0.5% SDS, $0.2 \cdot 10^6$ cpm/ml galactokinase mRNA were added to the bag. The bag was sealed and incubated at 41°C for 24 h.

After hybridization the filters were washed twice with 50% formamide in 4 × SSC for 20 min with 200 ml each time at room temperature, and three to four times with 2 × SSC, again for 20 min each. The filters were dried and autoradiographed for 1 week.

Hybridization analysis of putative galactokinase containing plasmids

To prepare plasmid DNA fixed to nitrocellulose filters (Cooper et al., 1974), 50 ng of plasmid DNA was added to 0.5 ml of 0.01 × SSC containing 20 μg/ml *E. coli* DNA. The DNA was denatured by heating at 98°C for 12 min and quickly chilled to 0°C. The solution was adjusted to 6 × SSC in a final volume of 2.0 ml, passed slowly through 25 mm nitrocellulose filters; the filters were washed and then baked. From each filter 5 mm diameter circles were punched out with a standard paper punch; each of these filters contained 3–4 ng of plasmid DNA.

Competition hybridization of the ^{125}I-labelled galactokinase mRNA to these immobilized plasmid DNAs was performed by placing each disc in 0.05 ml of hybridization solution (50% formamide, 0.8 M NaCl, 0.5% SDS, 50 mM Tris·HCl pH 7.5, 10 mM EDTA, 260 μg/ml induced or non-induced polysomal mRNA) in conical microfuge tubes. After pre-hybridization for 40 h at 41°C, 10 ng of ^{125}I-labelled galactokinase mRNA ($0.2 \cdot 10^6$ cpm) in 0.01 ml of hybridization solution were added to each hybridization tube. After an additional incubation at 41°C for 40 h, the nitrocellulose filters were assayed for hybrid formation by a modification of the method of Cooper et al. (1974).

Assay of galactokinase activity

Galactokinase assays were performed by a modification of the method of Schell and Wilson (1977) using [^{14}C]galactose with a specific activity of 10 000 cpm/nmole. 10 mM glucose was included in some assays to block phosphorylation of galactose by enzymes other than yeast galactokinase. Protein was measured by the method of Lowry et al. (1951) using bovine serum albumin for reference.

Preparation of mRNA

Total polysomal mRNA from galactose-induced or uninduced yeast was prepared from *S. cerevisiae* strain X108D as described by Schell and Wilson (1979). Total induced or uninduced mRNA was prepared by deproteinization

of a post-mitochondrial supernatant extract of the same cells with chloroform: phenol:isoamyl alcohol and chromatography on polyuridylic acid-linked sepharose.

Preparation of media

M9 medium was prepared by adding the following chemicals to 1.0 liter of distilled water: 6 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl, 1 g NH_4Cl. After autoclaving the medium was made to 1 mM $MgSO_4$ and 0.1 mM $CaCl_2$ by the addition of appropriate amounts of sterile 1.0 M stock solutions. L-broth was prepared by adding 10 g of bacto-tryptone, 5 g NaCl, and 5 g yeast extract to 1.0 liter of water; plates contained 2.0% agar. Carbon sources were included at a final concentration of 0.2%.

Chemicals

$TlCl_3$ (K and K Laboratories); $Na^{125}I$ (New England Nuclear; carrier free); KI (Mallinckrodt); Agarose (Seakem); Nitrocellulose sheets (Millipore); CsCl (Kawecki-Berylco); phenol (Mallinckrodt; distilled and stored at −20°C); formamide (Fisher; distilled in vacuo and stored at −20°C); DNA enzymes: *Bam*HI, *Bgl*II, *Eco*RI, T4 ligase (New England Biolabs); ampicillin (Parke-Davis; Amsil-S); tetracycline (Sigma). All other chemicals were purchased from Sigma Chemical Co. or were of reagent grade purity.

Recombinant DNA procedures

These experiments were carried out in strict compliance with the NIH Guidelines for research involving recombinant DNA molecules. The level of containment employed was P2 EK1 as prescribed by these guidelines.

RESULTS

Identification of a restriction fragment containing the galactokinase gene

Yeast DNA was digested with *Bgl*II restriction enzyme, subjected to agarose gel electrophoresis, and the resulting pattern of DNA fragments transferred to nitrocellulose filters as described in METHODS. These DNA filters were hybridized with ^{125}I-labelled galactokinase mRNA and autoradiographed. Three discrete bands (A,B,C) were consistently detected in the *Bgl*II restriction enzyme digest. This is illustrated in Fig. 1 (slot 2) which is a typical autoradiograph of a nitrocellulose sheet containing a *Bgl*II digest hybridized with galactokinase mRNA. Due to the larger amount of DNA coding for rRNA relative to single copy genes it was likely that at least one of these bands resulted from the hybridization of small amounts of rRNA which contaminates most RNA preparations. This conclusion was supported by preliminary Southern filter hybridization experiments with *Eco*RI digests of yeast DNA, where the galactokinase mRNA preparation hybridized to known rDNA restriction enzyme fragments (Bell et al., 1977). Furthermore fragment C in the hybridization pattern in Fig. 1 has a molecular weight which is

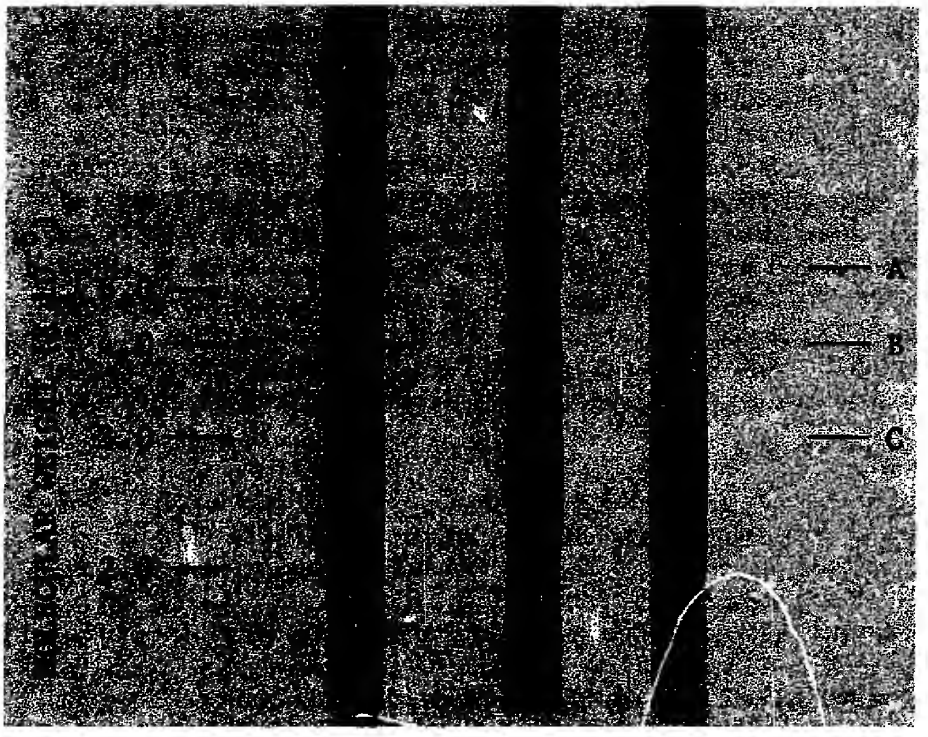


Fig. 1. Hybridization of ^{125}I-labelled galactokinase mRNA to *Bgl*II restriction enzyme digests of yeast DNA. Yeast DNA was digested with *Bgl*II restriction enzyme, fractionated by agarose gel electrophoresis and transferred to nitrocellulose sheets. The DNA digests were hybridized with: Slot 1 = 10 μg of *Bgl*II digested DNA hybridized with 10^6 cpm of ^{125}I-labelled total induced polysomal mRNA. Slot 2 = 10 μg of *Bgl*II digested DNA hybridized with 10^6 cpm of ^{125}I-labelled galactokinase mRNA. Slot 3 = 10 μg of *Bgl*II digested DNA hybridized with 10^6 cpm of ^{125}I-labelled yeast ribosomal RNA. Molecular weight values were calculated using *Eco*RI digested λ-phage DNA that was run on the same gel as a reference.

identical to the size of the *Bgl*II restriction enzyme fragments containing the yeast rDNA (J. Szostak, personal communication).

This hypothesis was confirmed by hybridizing *Bgl*II digests of yeast DNA on nitrocellulose sheets with ^{125}I-labelled yeast rRNA or total mRNA from non-induced yeast cells. The results of these hybridizations are illustrated in Fig. 1 (slots 1 and 3). It is clear that fragments A and C are probably rDNA fragments since they are the only bands in *Bgl*II digests that hybridize yeast rRNA and non-induced mRNA. Fragment B did not hybridize to either rRNA or non-induced mRNA and therefore it was likely that fragment B with a molecular weight of $3.7 \cdot 10^6$ was the *Bgl*II restriction fragment containing the galactokinase gene.

Partial purification of the galactokinase gene fragment

In order to isolate preparative amounts of this DNA fragment, 400 μg of yeast DNA were digested with 30 units of *Bgl*II restriction enzyme for 22 h

at 37°C in the solution described in METHODS. After ethanol precipitation the redissolved DNA was run on a preparative 1% agarose gel (4 slots; 0.6 cm × 4 cm) at 70 mA for 18 h; λ DNA cut with *Eco*RI was run on the same gel as a molecular weight marker.

Since the molecular weight of the putative galactokinase gene containing *Bgl*II restriction fragment was $3.7 \cdot 10^6$, the area of the gel containing the restriction enzyme fragments of molecular weight $3.4 \cdot 10^6$ to $4.0 \cdot 10^6$ was cut from the gel and the DNA extracted from the agarose pieces as described in METHODS. The final yield of DNA from this partial purification procedure was 20—30 μg. Since 50% of the DNA was lost during manipulation, it can be estimated that this partially purified DNA fraction containing the galactokinase gene fragment had been enriched 10—20-fold for this specific fragment.

Generation of a clone bank

The DNA isolated from the gel was ligated into pBR322 plasmid DNA at the *Bam*HI restriction enzyme site as described in METHODS. The ligation appeared to be greater than 50% complete as estimated by electrophoresis using a 1% agarose gel in Tris·borate buffer (Greene et al., 1977). The mixture of ligated plasmid and plasmid with inserted yeast DNA fragments was used to transform *E. coli* strain M94 by the procedure described in METHODS. Selection for transformed cells containing pBR322 on plates containing 0.1 mg/ml ampicillin showed that $1.5 \cdot 10^3$ cells/ml were ampicillin-resistant out of a total of $1.7 \cdot 10^8$ viable cells/ml indicating a frequency of 10^{-5} transformation to ampicillin resistance per cell with the plasmid preparation. The entire transformation mixture was plated out on 25 ampicillin plates (0.1 ml = 150 colonies/plate). When these plates were replica plated on tetracycline plates approx. 15% of the total ampicillin-resistant transformants were found to be tetracycline-sensitive. Since the yeast DNA fragments are inserted into the tetracycline-resistance gene on the plasmid (Rodriguez et al., 1977), these 15% were presumed to contain plasmid with yeast DNA ligated into it. From three individual transformations 1500 ampicillin-resistant/tetracycline-sensitive transformants were isolated. It was calculated from the purification factor of the galactokinase gene fragment by agarose gel electrophoresis, that there should be from 300—800 unique DNA fragments in this clone bank and therefore it should contain from 2—5 plasmid clones with the *Bgl*II restriction enzyme fragment containing the galactokinase gene.

Detection of plasmid containing the galactokinase gene

Filters were prepared with DNA from each of the plasmid strains fixed at discrete locations on Whatman 540 paper as described in METHODS. After hybridization of these filters and autoradiography, only 6 colonies out of the 1500 transformants hybridized to ^{125}I-labelled galactokinase mRNA. Each of these strains were grown and plasmid DNA isolated as described by Meagher et al. (1977). After digestion of these plasmids with *Eco*RI and gel electrophoresis in 1% agarose none of the six plasmids appeared to have

identical band patterns. Furthermore, two of the six unique plasmids appeared to contain a piece of yeast DNA that was too small to be the $3.7 \cdot 10^6$ *Bgl*II fragment. The origin of these small plasmids is not known although cellular modification of the plasmid is a possible explanation. However four of the plasmids contained yeast DNA fragments which might contain the galactokinase gene.

Since it has been shown that mRNA isolated from non-induced polysomes contained no galactokinase mRNA (Schell and Wilson, 1979), the plasmid containing the yeast galactokinase gene could be identified by competition hybridization. Competition hybridizations of the ^{125}I-labelled galactokinase mRNA to the various plasmid DNAs fixed on nitrocellulose filters were performed after pre-hybridization with an excess of unlabelled mRNA from both induced and non-induced yeast cells as described in METHODS. The results are presented in Table I. As expected the hybridization of the ^{125}I galactokinase mRNA to the various plasmid DNAs pre-hybridized with an excess (12 μg mRNA/2 ng plasmid) of unlabelled induced mRNA is completely blocked. Hybridization of two of the plasmid DNAs (GB-4 and GB-7) was not blocked by an excess of non-induced mRNA while the rest of the plasmids were totally blocked. We conclude from this experiment that the plasmids designated GB-4 and GB-7 contain a sequence homologous to a galactose induced mRNA since the hybridization of galactokinase mRNA to them cannot be blocked by unlabelled non-induced (i.e. non galactokinase-containing) mRNA. Furthermore, when these plasmids (labelled by nick translation [Rigby et al., 1977]) were hybridized to *Bgl*II digests of whole yeast DNA on nitrocellulose filters both plasmids hybridized to the $3.7 \cdot 10^6$ dalton restriction fragment that

TABLE I

HYBRIDIZATION OF GALACTOKINASE mRNA TO DIFFERENT PLASMID DNAs PRE-HYBRIDIZED WITH VARIOUS mRNA FRACTIONS

Galactokinase mRNA (10 ng = $0.5 \cdot 10^6$ cpm ^{125}I) was hybridized to nitrocellulose filters containing 3 ng of the various plasmid DNAs as described in METHODS. The filters were first pre-hybridized with no unlabelled mRNA, 0.012 mg of unlabelled polysomal mRNA from non-induced yeast cells (+ NI mRNA), or 0.012 mg of polysomal mRNA from galactose induced yeast cells (+ I mRNA). Values are expressed both as the ^{125}I cpm of galactokinase mRNA hybridized or as the percentage of hybridization in the absence of pre-hybridization with unlabelled mRNA.

Plasmid designation	^{125}I cpm hybridized		
	No pre-hybridization	+ NI mRNA (%)	+ I mRNA (%)
GB-2	830	10 (1)	20 (2)
GB-3	2120	23 (2)	15 (1)
GB-4	860	840 (98)	25 (3)
GB-5	5100	50 (1)	60 (1)
GB-6	750	27 (4)	20 (3)
GB-7	350	325 (93)	15 (5)

hybridized only galactokinase mRNA shown in Fig. 1. The plasmid GB-7 also contained an additional fragment of slightly smaller molecular weight; this fragment was "co-cloned" into a plasmid with the galactokinase fragment to produce the larger plasmid. This is supported by the fact that GB-4 DNA was resistant to *Bgl*II digestion, while GB-7 was cleaved by *Bgl*II to produce a linear DNA molecule. Thus GB-4 and GB-7 are the same plasmid except that GB-7 contains an additional *Bgl*II fragment of DNA derived from an insertion of another individual DNA fragment into the plasmid molecule.

Expression of yeast galactokinase gene in E. coli

Proof that the cloned fragment, which is homologous to a galactose-induced mRNA, does contain the functional galactokinase gene was provided by the following experiment. The GB-4 plasmid DNA was used to transform *E. coli* strain S165 which contains a deletion of the *E. coli* galactose operon (Starlinger, pers. commun.). When extracts of this strain containing the GB-4 plasmid were analysed for galactokinase activity, a low but significant activity (Table II) was detected. Extracts of this strain containing the GB-4 plasmid exhibited a galactokinase activity of $2 \cdot 10^{-3}$ U./mg, while extracts of the same strain containing no plasmid or the pBR322 plasmid have an apparent activity of $1 \cdot 10^{-3}$ U./mg. When the same assays were performed in the presence of 10 mM glucose to block phosphorylation of galactose not catalysed by yeast galactokinase (Schell and Wilson, 1977), the GB-4 plasmid bearing strain still had an activity of $1 \cdot 10^{-3}$ U./mg while the control strains (with pBR322 only or with no plasmid) have galactokinase activity which is at the limit of detection of the galactokinase assay procedure. This experiment shows that the cloned fragment contains a functional yeast galactokinase gene, which can be expressed at a low level in *E. coli*.

TABLE II

EXPRESSION OF GALACTOKINASE ACTIVITY IN VARIOUS PLASMID-CONTAINING STRAINS

Cells of *E. coli* strain S165 containing the designated plasmids were grown in 10 ml of L-broth containing 0.2% galactose and 0.1 mg/ml ampicillin to $A_{600\,nm} = 0.8$. The cells were harvested, washed, and cell extracts prepared by sonication. Extracts were then assayed for galactokinase activity in the presence and absence of 10 mM glucose in the assay mix. Activity levels in these strains is expressed in the units/mg protein · 10^3.

Extract source	Galactokinase activity (U./mg · 10^3)	
	no glucose	+ 10 mM glucose
S165 + 0	1.0 ± 10%	0.05 ± 20%
S165 + pBR322	1.0 ± 10%	0.05 ± 20%
S165 + GB-4	2.0 ± 10%	1.0 ± 10%

DISCUSSION

The isolation of a plasmid containing the yeast galactokinase gene in an *E. coli* host is the final step in the purification of the components of the expression of the galactokinase gene. The successful isolation of this plasmid was dependent on the sequential purification of galactokinase (Schell and Wilson, 1977), galactokinase antibody, and galactokinase mRNA (Schell and Wilson, 1979) since each of these purified components was necessary to produce an assay system for the galactokinase gene. This successful cloning experiment provides support for the general use of this sequential isolation method (protein → antibody → mRNA → DNA) to clone and study many other eukaryotic genes and their components. Although the methods used to isolate the galactokinase gene are not new, the cloning of a specific gene whose mRNA represents less than 2% of the total cellular mRNA utilizing these methods has not been reported.

The isolation of the galactokinase plasmid was dependent on the use of very high specific activity ^{125}I-labelled purified galactokinase mRNA to specifically hybridize to the galactokinase gene fragment in *Bgl*II digests of the entire yeast genome, allowing subsequent partial purification of the fragment 10- to 20-fold. Although other restriction enzymes were investigated, the enzyme *Bgl*II was found to produce the largest DNA fragment ($3.7 \cdot 10^6$) containing the galactokinase gene which was easily separated from ribosomal DNA. Due to its large size there was a possibility that this DNA fragment would also contain other parts of the galactose gene cluster in addition to the galactokinase gene.

The partially purified DNA preparation was ligated in the *Bam*HI site of the plasmid pBR322 inserting the yeast DNA into the plasmid gene for tetracycline resistance. This allowed elimination of the large number of transformants which contain re-ligated plasmid without inserted yeast DNA (85%). In situ colony hybridization to 1500 of these clones with ^{125}I-labelled galactokinase mRNA detected six individual plasmids that strongly hybridized the galactokinase mRNA preparation. However, for four of the six plasmids this hybridization could be completely blocked by an excess of unlabelled mRNA from cells that were not induced for the synthesis of galactokinase. Two different plasmids (GB-4 and GB-7) hybridized galactokinase mRNA in the presence of the same amount of non-induced mRNA that completely blocked the hybridization by the other plasmids. It can be concluded from these results that these two plasmids contain a sequence coding for a galactose inducible mRNA. The identity of the other four plasmids that hybridize galactokinase mRNA but are blocked by an excess of non-induced mRNA is not known. They could contain sequences of DNA that are homologous to the non-coding portion of the galactokinase mRNA, or represent DNAs homologous to contaminating mRNAs in the galactokinase mRNA preparation.

Hybridization experiments with the galactokinase plasmids showed that a large excess of total cellular mRNA or total polysomal mRNA from uninduced

cells does not block the hybridization of galactokinase mRNA to the plasmid. This suggests that there is an undetectable amount of galactokinase mRNA in yeast cells that are not exposed to galactose. Therefore it seems that the induction of galactokinase synthesis (and probably the entire galactose gene cluster) by galactose is the result of the induction of the synthesis of mRNA's coding for these proteins. Transcriptional control therefore appears to be the predominant mechanism by which the induction of galactose enzyme synthesis is mediated.

The proof that one of the isolated plasmids contains a functional yeast galactokinase gene was provided by the experiments that demonstrated the expression of the yeast galactokinase gene in *E. coli*. An *E. coli* strain, deleted for its galactose operon, carrying the galactokinase plasmid exhibited a galactokinase activity of 0.001 U./mg when assayed under conditions that are specific for yeast galactokinase. Galactokinase assays of the same strain containing either no plasmid or only pBR322 plasmid showed no apparent galactokinase activity under the same conditions. Although this level of expression of the yeast galactokinase gene in *E. coli* is low, the activity is real and therefore strongly suggests that the isolated plasmid does indeed contain a functional yeast galactokinase gene. Therefore the yeast galactokinase gene can be added to the list of yeast genes which have been cloned and shown to be expressed in *E. coli* (Clarke and Carbon, 1978).

ACKNOWLEDGEMENTS

The authors would like to thank Drs. J. Calvo, R. Wu, R. Rothstein, S. Wessler, and J. Szostak for their advice, comments, and assistance.

This research was supported in part by Grant PCM 77-15487 from the National Science Foundation.

REFERENCES

Beckmann, J.S., Johnson, P.F. and Abelson, J., Cloning of yeast transfer RNA genes in *Escherichia coli*, Science, 196 (1977) 205–208.

Bell, G.I., DeGennaro, L.J., Gelfand, D.H., Bishop, R.J., Valenzuela, P. and Rutter, W.J., Ribosomal genes of *Saccharomyces cerevisiae*, I. Physical map of the repeating unit and location of the regions coding for 5S, 5.8S, 18S, and 25S ribosomal RNAs, J. Biol. Chem., 252 (1977) 8118–8125.

Burton, K., Determination of DNA concentration with diphenylamine, in Grossman, L. and Moldave, K. (Eds.), Methods in Enzymology, Vol. XIIB, Academic Press, New York, 1968, pp. 163–166.

Clarke, L. and Carbon, J., Expression of cloned yeast DNA in *Escherichia coli*: specific complementation of argininosuccinate lyase (*argH*) mutations, J. Mol. Biol., 120 (1978) 517–532.

Cooper, T.G., Whitney, P. and Magasanik, B., Reaction of *lac*-specific ribonucleic acid from *Escherichia coli* with *lac* deoxyribonucleic acid, J. Biol. Chem., 249 (1974) 6548–6555.

Cryer, D., Eccleshall, R. and Marmur, J., Isolation of yeast DNA in Prescott, D. (Ed.), Methods in Cell Biology, Vol. XII, Chapter 3, Academic Press, New York, 1975, pp. 39–44.

Greene, P.M., Betlach, M., Goodman, J., and Boyer, H.W., Gel electrophoresis of plasmid DNA, in Wickner, R.B. (Ed.), Methods in Molecular Biology, Vol. 7, Marcel Dekker, New York, 1974, pp. 87—111.

Lowry, O.H., Rosebrough, N.J., Farr, A.L. and Randall, R.J., Protein measurement with the Folin phenol reagent, J. Biol. Chem., 193 (1951) 265—275.

Meagher, R.B., Tait, R.C., Betlach, M. and Boyer, H.W., Protein expression in *E. coli* minicells by recombinant plasmids, Cell, 10 (1977) 521—536.

Miller, J.H., Replica plating, in Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, New York, 1972, pp. 56—59.

Morrison, D.A., Transformation in *Escherichia coli*: cryogenic preservation of competent cells, J. Bacteriol., 132 (1977) 349—351.

Prensky, W., Radiodination of DNA and RNA in Prescott, D. (Ed.), Methods in Cell Biology, Vol. XIII, Chapter 7, Academic Press, New York, 1976, pp. 122—152.

Rigby, P.W.J., Dreckmann, M., Rhodes, C. and Berg, P., Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I, J. Mol. Biol., 113 (1977) 237—251.

Rodriguez, R.L., Tait, R., Shine, J., Bolivar, F., Heyneker, H., Betlach, M. and Boyer, H.W., Characterization of tetracycline and ampicillin resistant plasmid cloning vehicles, in Scott, W.A. and Werner, R. (Eds.), Molecular Cloning of Recombinant DNA (Miami Winter Symposia; Vol. 13) Academic Press, New York, 1977, pp. 73—84.

Schell, M.A. and Wilson, D.B., Purification and properties of galactokinase from *Saccharomyces cerevisiae*, J. Biol. Chem., 252 (1977) 1162—1166.

Schell, M.A., and Wilson, D.B., Purification of galactokinase mRNA from *S. cerevisiae* by indirect immunoprecipitation, J. Biol. Chem. (1979) in press.

Southern, E.M., Detection of specific sequences among DNA fragments separated by gel electrophoresis, J. Mol. Biol., 98 (1975) 503—517.

Tanaka, T. and Weisblum, B., Construction of a colicin E1-R factor composite plasmid in vitro: means for amplification of deoxyribonucleic acid, J. Bacteriol., 121 (1975) 354—362.

Wu, R., Jay, E., and Roychoudhury, R., Nucleotide sequence analysis of DNA, in Busch, H. (Ed.), Methods in Cancer Research, Vol. XII, Chapter III, Academic Press, New York, 1976, pp. 88—177.

Communicated by F.E. Young.

TAB R

Proc. Natl. Acad. Sci. USA
Vol. 76, No. 1, pp. 386–390, January 1979
Genetics

Evidence for transcriptional regulation of orotidine-5′-phosphate decarboxylase in yeast by hybridization of mRNA to the yeast structural gene cloned in *Escherichia coli*

(complementation/*in vitro* recombination/*Saccharomyces cerevisiae*/pyrimidine biosynthesis)

MARIE-LOUISE BACH*†, FRANCOIS LACROUTE†, AND DAVID BOTSTEIN*‡

*Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; and †Laboratoire de Genetique Physiologique, Institut de Biologie Moléculaire et Cellulaire du Centre National de la Recherche Scientifique, 67084 Strasbourg, France

Communicated by Boris Magasanik, August 18, 1978

ABSTRACT From a large population of strains of *Escherichia coli* carrying shear fragments of yeast (*Saccharomyces cerevisiae*) DNA attached by *in vitro* recombination to the plasmid vector pMB9, two hybrid plasmids were selected that relieve the pyrimidine requirement of nonreverting *pyrF* mutants of *E. coli*. An 1100-base-pair DNA fragment common to the two complementing plasmids was recloned into another plasmid vector, pBR322; these new hybrids retained the ability to specify orotidine-5′-phosphate decarboxylase (orotidine-5′-phosphate carboxy-lyase, EC 4.1.1.23) synthesis in *E. coli*. Evidence is presented that this common fragment is yeast DNA and thus apparently carries the structural information for yeast orotidine-5′-phosphate decarboxylase, the product of yeast gene *ura*3. A hybrid plasmid containing the 1100-base-pair fragment was used to measure levels of putative *ura*3 mRNA from yeast cultures labeled with [^{3}H]adenine. *ura*3 mRNA was unstable with an apparent half-life of 10.5 min. Under different circumstances previously shown to alter the level of orotidine-5′-phosphate decarboxylase in yeast, a coordinate variation in proportion of labeled RNA complementary to the hybrid plasmid was found. These data support the hypothesis that regulation of the *ura*3 gene in yeast is at the level of transcription.

The recent development of *in vitro* recombination methods and the associated technology has made it possible to isolate specific eukaryotic genes. In the case of yeast (*Saccharomyces cerevisiae*) it has been found that several genes for enzymes in metabolism can be expressed when introduced into *Escherichia coli* on either plasmid or phage vehicles (1–3). We report in this paper that the yeast gene that specifies orotidine-5′-phosphate (OMP) decarboxylase (orotidine-5′-phosphate carboxy-lyase, EC 4.1.1.23), the last enzyme in the pathway for the biosynthesis of pyrimidines, can be expressed in *E. coli*. The gene's complementation activity was retained by a fragment of DNA about 1100 base pairs long.

From genetic and physiological studies of the regulation of the pyrimidine pathway in yeast, Lacroute (4) had obtained evidence that the *ura*3 gene (which encodes OMP decarboxylase) is inducible by earlier intermediates in the pyrimidine pathway. For example, he found that mutants in another gene (*ura*1; defective in dihydroorotase), which accumulate high intracellular levels of dihydroorotic acid, contain 5 times more OMP decarboxylase.

We used a hybrid plasmid containing the 1100-base-pair yeast DNA fragment containing the *ura*3 gene to measure levels of *ura*3 mRNA in yeast by hybridization. We found that the level of *ura*3 mRNA varied coordinately with the levels of the OMP decarboxylase activity and concluded that the *ura*3 gene is regulated at the level of transcription.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Yeast and Bacteria. Pyrimidine auxotrophs of *S. cerevisiae* were all derived from the wild-type prototroph *a* strain FL100 (4, 5). The construction of the yeast/plasmid hybrids is fully described elsewhere (6). All manipulations involving live *in vitro* clones were done under P2,EK1 containment as suggested by the National Institutes of Health guidelines (7). The bacterial strains all were derivatives of *E. coli* K-12. Strain DB6656 is *pyrF::Mu trp*$_{am}$ *lacZ*$_{am}$ *hsdR*. The pyr$^-$ mutation reverts at less than 10^{-9} and was confirmed by enzyme assay. The plasmids were pMB9 (*tet*^{R}) and pBR322 (*tet*^{R} *amp*^{R}) (8).

Screening of DNA Pools for Complementation. In order to make DNA pools, the hybrid yeast/pMB9 clones were inoculated onto LB (9) agar plates containing 15 μg of tetracycline per ml and incubated overnight at 37°C. The cells were washed off these plates into erlenmeyer flasks (about 250 clones per flask) containing LB broth with tetracycline. The total growth of the pooled clones together before DNA extraction was only about 2 doublings, ensuring that each clone that grew up into a spot on the plate was well represented in the extracted plasmid DNA. The DNA pools were used for transformation of strain DB6656 (6, 10), selecting for pyrimidine independence.

Analysis of Complementing Hybrid Plasmids. Isolation of plasmid DNA and retransformations were done exactly as described (6). Analysis of plasmid DNA with site-specific restriction endonucleases was carried out as described by Sharp *et al.* (11). The hybridization of "nick-translated" radioactive plasmid DNA (12) to digested yeast DNA by the method of Southern (13) was carried out essentially as described (14, 15). The slots containing plasmid DNA were loaded with about 1 μg of DNA whereas the yeast slots contained 5 μg each; this accounts for the immense imbalance in degree of hybridization.

The cloning of *Hin*dIII fragment b into pBR322 was accomplished by standard methods (16, 17). The ligated DNA was used to transform strain DB6656, selecting for growth on minimal plates lacking uracil. Transformants were purified and tested for resistance to ampicillin and tetracycline. Some of the *amp*^{R} *Pyr*$^+$ strains were more resistant to tetracycline than others, but all failed to grow in the presence of tetracycline (15 μg/ml), consistent with the idea that the b fragment had been inserted into the *Hin*dIII site of pBR322 (8).

Determination of OMP Decarboxylase Activity. The method of Wolcott and Ross (18) was used, with minor modifications. Extracts of yeast and bacteria were made by sonication in the cold. Specific activity was determined relative to the protein concentrations (19).

Labeling and Extraction of Yeast RNA. Labeling of RNA

Abbreviation: OMP, orotidine 5′-phosphate.

‡ To whom requests for reprints should be addressed.

was done with [^{3}H]adenine (20 μCi/ml; 25 Ci/mmol; Saclay), which was added to yeast cells growing exponentially in excess uracil. Samples were taken and the cells were either arrested by the addition of 2 vol of cold ethanol or chased with nonradioactive adenine (50 μg/ml) and arrested later. During the chase, total acid-insoluble radioactivity increased no more than 40% in the first few minutes and remained constant thereafter. Equilibrium labeling was done by incubation for at least three generations in medium containing 20 μCi of [^{3}H]adenine per ml in a total adenine concentration of 10 μg/ml; these conditions result in only 20% incorporation of the radioactivity.

RNA was extracted from cells by a minor modification of methods described previously (20); no carrier RNA was added and sodium dodecyl sulfate was omitted from the second ethanol precipitation. RNA samples were lyophilized after the final ethanol precipitation and taken up in hybridization buffer.

Hybridization. Nitrocellulose filters with plasmid DNA were prepared by filtration under denaturing conditions (21, 22). Sartorius filters (type 11306, 0.45 μm, 25 mm diameter) were loaded with plasmid DNA, and 5-mm-diameter microfilters containing 1–2 μg of DNA were punched out of them. Hybridization was done under conditions slightly modified from those of Kourilsky *et al.* (23). The hybridization solution was 0.3 M NaCl/0.03 M Na citrate/40% (vol/vol) formamide containing 0.02% bovine serum albumin, 0.02% polyvinylpyrrolidone, 0.02% Ficoll, and 0.2% sodium dodecyl sulfate to decrease the background (24). Radioactive RNA was dissolved in this solution (usually 2 ml) and 0.2-ml aliquots were incubated for 2 days with the DNA filters at 37°C in a sealed glass vial. Vials contained filters loaded with yeast hybrid plasmid DNA and, as control, with vector plasmid DNA. Control experiments verified that hybridization was complete under these conditions when there was 0.3 μg or more of DNA on the filter and that even with the most abundant RNA there was a 20-fold excess of DNA on the filters. Filters were washed, treated with RNase, dried, and assayed for radioactivity in a scintillation counter for at least 20 min.

Hybridizations were done in triplicate or quadruplicate. Values given in tables and figures are the difference between two means—(mean cpm on filters loaded with yeast/vector hybrid DNA) − (mean cpm on filters loaded with vector DNA)—divided by the total acid-insoluble radioactivity. The background was about 40 cpm (filters with vector DNA gave results not significantly different from filters with no DNA); usually more than 10^6 cpm of acid-insoluble radioactivity was applied.

RESULTS

Isolation of Hybrid Plasmids in *E. coli* Carrying the Yeast Gene Specifying OMP Decarboxylase. Selection for function of the yeast gene specifying OMP decarboxylase was performed by using a nonreverting mutation in the *pyrF* gene, which is the *E. coli* structural gene for OMP decarboxylase (24). A set of 2500 yeast/*E. coli* plasmids was constructed by insertion of randomly sheared pieces of *S. cerevisiae* DNA into the small *E. coli* plasmid pMB9 by the terminal-transferase method. The details of the construction of this set were described by Petes *et al.* (6); the probability that any particular yeast DNA sequence is present in this set was estimated to be 0.85. Groups of 250 *E. coli* strains, each of which harbored a hybrid plasmid, were grown up and pooled, and the plasmid DNA was extracted by standard methods (25). Ten pools of yeast/*E. coli* plasmid DNA were used to transform the nonreverting *E. coli pyrF::Mu* recipient. After a short period of growth in complete medium, the transformed bacteria were spread on agar plates containing no pyrimidines. Two of the 10 plasmid DNA pools yielded some Pyr$^+$ transformants; the 8 others yielded none.

If the Pyr$^+$ phenotype is the result of the presence of a specific yeast DNA sequence, then the property should be transferable by transformation with plasmid DNA, whether or not selection for growth in the absence of pyrimidines is carried out. This was confirmed in two ways. First, the DNA from Pyr$^+$ transformants was used to transform again the *pyrF::Mu* strain, selecting for tetracycline resistance. All drug-resistant transformants acquired simultaneously the ability to grow in medium lacking pyrimidines. Second, another *E. coli* strain (HB101) which does not require pyrimidines was transformed to tetracycline resistance. Plasmid DNA was extracted from these transformants and tested for ability to transform the *pyrF::Mu* strain to Pyr$^+$. Again, all drug-resistant transformants yielded DNA with the ability to confer Pyr$^+$ phenotype. Two hybrid plasmids, each deriving from one of the two successful DNA pools, were selected for further study; these are called clone 1 and clone 2.

If the complementation of the *pyrF* defect by the DNA of clones 1 and 2 were due to the presence of yeast DNA specifying OMP decarboxylase, then the two plasmids should contain common sequences derived from yeast. In order to test this further, analysis of the two plasmid DNAs was carried out by using site-specific restriction endonucleases. The first results indicated that the two plasmids were different in length, and that neither was cut by *Eco*RI. However, both were cut several times by *Hin*dIII, yielding two fragments of identical mobility as judged by electrophoresis in agarose slabs (Fig. 1). The vector (pMB9) has only one *Hin*dIII site. Differences in length between the two plasmids were expected because the original hybrids were made from randomly sheared yeast DNA, and the endonuclease digestion results are consistent with the idea that the two plasmids are independent isolates of the same region of yeast DNA.

In order to try to determine which parts of the two plasmids contained the DNA responsible for the complementation of the *pyrF* defect, an experiment was carried out in which the *Hin*dIII fragments were ligated into another *E. coli* plasmid vector, pBR322 (8). This vector carries genes conferring resistance to both ampicillin and tetracycline. The insertion of DNA into the single *Hin*dIII site on this vector causes loss of the tetracycline resistance (8). *Hin*dIII digests of clones 1 and 2 were mixed with *Hin*dIII-digested pBR322 DNA and treated with T4 polynucleotide ligase. The ligated DNA was used to transform *E. coli pyrF::Mu* and, as before, selection was made for independence of pyrimidines. Pyr$^+$ colonies were tested for resistance to

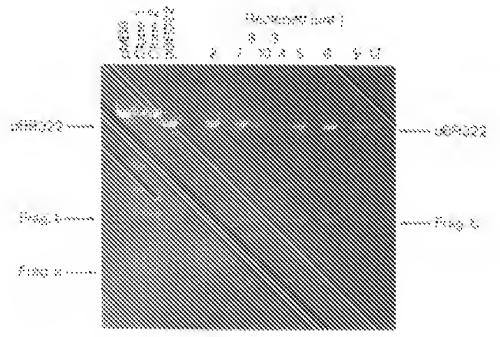


FIG. 1. Agarose/ethidium bromide electropherograms of plasmid DNAs. Samples of purified plasmid DNA (approx. 0.5 μg per lane) were digested with *Hin*dIII and electrophoresed as described by Sharp *et al.* (11). Lanes labeled "Recloned" are plasmids obtained by recloning the *Hin*dIII digests of clones 1 and 2 into pBR322 followed by selection of Pyr$^+$ after transformation of strain DB6656.

ampicillin and tetracycline; Amp^R, Tet^S, Pyr^+ colonies were found, purified by single-colony isolation, and grown up, and plasmid DNA was isolated from them. After digestion with *Hin*dIII, each of these plasmids yielded two fragments, one with the mobility of pBR322, and one with the mobility of the larger of the two *Hin*dIII fragments (fragment b) common to clones 1 and 2 (Fig. 1). Thus, the ability to confer Pyr^+ phenotype could be narrowed down to a single fragment. Analysis of the length of this fragment in gels with ϕX174 markers gave a result of about 1100 base pairs (data not shown).

It was important to show that the complementing DNA in the hybrid plasmids was derived from yeast. The technique of Southern (13) in which DNA fragments displayed on an agarose gel are adsorbed to nitrocellulose paper and hybridized with labeled probes is ideal for this purpose. Several such experiments were carried out on *Eco*RI and *Hin*dIII digests of total yeast DNA from two different strains of yeast, using as probe the hybrid plasmids described above. Fig. 2 shows one such experiment. The radioactive probe was clone 6 DNA made radioactive by nick translation (12). Hybridization was observed at one position in the *Hin*dIII digest of yeast DNA corresponding to the size of fragment b and at one position in the *Eco*RI digest of yeast DNA corresponding to a fragment larger than the sum of all the yeast *Hin*dIII fragments of clones 1 and 2. This result demonstrates that fragment b is a fragment of yeast DNA and that yeast DNA is responsible for the complementation of the *pyrF* mutation in *E. coli*.

Other Southern-type experiments, as well as standard restriction enzyme analysis (not shown), permit the unambiguous construction of a restriction map of the region of the yeast genome from which clones 1 and 2 derive (Fig. 3). These results

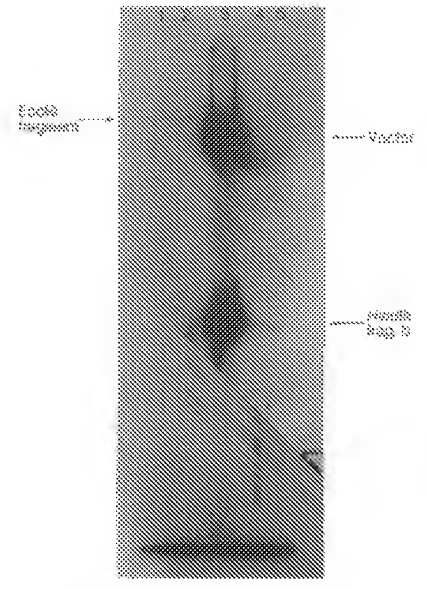


FIG. 2. Hybridization of clone 6 DNA to total yeast DNA. Total yeast DNA (5 μg) from strains FL100 (lanes 1 and 4) and +D4 (lanes 2 and 5) was digested with *Eco*RI (lanes 1 and 2) or *Hin*dIII (lanes 4 and 5). Lane 3 contained 1 μg of *Hin*dIII-digested clone 1 DNA. The digested DNAs were electrophoresed through 1% agarose and transferred to nitrocellulose paper (13–15). "Nick-translated" (15) clone 6 DNA was hybridized to the paper. The bands in lanes 1 and 2 indicate that clone 6 hybridizes to the same single *Eco*RI fragment from both yeast strains and the bands in lanes 4 and 5 show that clone 6 hybridizes to a single *Hin*dIII band from both strands having the size of fragment b.

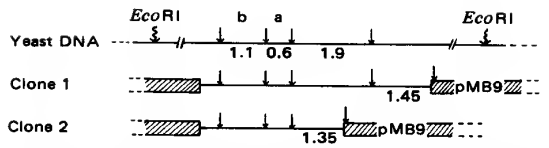


FIG. 3. Restriction map of yeast DNA in the neighborhood of the OMP decarboxylase gene. Straight arrows, *Hin*dIII cleavage sites; horizonal lines, yeast DNA; crosshatched bars, pMB9 vector plasmid DNA; numbers, approximate lengths in kilobases. The data for identity of fragments a and b in clones 1 and 2 are from Fig. 1. Partial digests and Southern experiments like Fig. 2 yielded the map.

show that the manipulations involved in constructing clones 1 and 2 did not cause any gross rearrangement of the DNA from yeast and that two widely differing yeast strains to whose DNA the probe was hybridized (+D4, from which the clones derive, and FL100, in which the *ura* genetics was done) are probably identical in this region of the genome.

The observation that plasmids bearing *Hin*dIII fragment b from yeast will allow a *pyrF* mutant of *E. coli* to grow in the absence of added pyrimidine makes it likely that fragment b contains the information for the synthesis of yeast OMP decarboxylase, because the *pyrF* mutation results in the loss of only this enzyme. The gene that specifies OMP decarboxylase in yeast (4) is called *ura*3, and it thus appears that fragment b bears the *ura*3 gene. Two lines of evidence support this idea. First, extracts of *E. coli pyrF::Mu* bearing clone 1, 2, or 6 contained substantial levels of OMP decarboxylase (Table 1). Second, preliminary results with the transformation procedure for yeast recently described by Hinnen *et al.* (26) indicate that clone 6 DNA will transform to uracil-independence a nonreverting *ura*3 mutant of yeast (M. Rose and D. Botstein, unpublished results).

Although clearly showing the presence of OMP decarboxylase activity resulting from the presence of the hybrid plasmids, the data in Table 1 also show that the level of enzyme activity somehow varies with conditions of growth and that it depends upon the plasmid used. These variations are not yet understood but might indicate some kind of modulation of gene expression even in *E. coli*.

For the present purpose, it suffices that fragment b is likely to contain all of the structural gene for yeast OMP decarboxylase and, from its small size (1100 base pairs), little else. Preliminary sequence information (D. Botstein and R. Tizard, unpublished data) is consistent with the idea that most of the 1100 base pairs code for a protein.

Regulation of Yeast OMP Decarboxylase Occurs at the Level of Transcription. Yeast OMP decarboxylase has been shown to be subject to induction by high intracellular levels of

Table 1. OMP decarboxylase activity in *E. coli*

Host	Plasmid	Growth conditions: Excess uracil	Growth conditions: Limiting conditions
E. coli pyrF	None	<0.001	<0.001
	Clone 1	0.10	3.2
	Clone 2	0.04	1.1
	Clone 6	5.9	12.6
E. coli pyrF$^+$	None	9.4	42.7

Cells were grown in M9 medium containing either excess uracil (40 μg/ml) or limiting uracil (1 μg/ml). When the cells had reached midexponential phase (excess uracil) or had starved for at least 1 hr (limiting uracil), they were harvested by centrifugation, resuspended in assay buffer, sonicated in the cold, and assayed for OMP decarboxylase. Values are given as mmol of OMP destroyed per mg of protein per minute.

dihydroorotic acid (and possibly by orotic acid and OMP as well). In *ura*1 mutants (defective in dihydroorotase), dihydroorotic acid accumulates. One finds in such mutants a level of OMP decarboxylase about 5 times greater than that in wild type; this level is only slightly affected if uracil is provided in excess (4). The best hypothesis for this induction is that enzyme biosynthesis is regulated, although alternative hypotheses have been difficult to rule out absolutely.

The isolation described above of the DNA for the structural gene of OMP decarboxylase (i.e., fragment b in pBR322, clone 6) made it possible to measure the level of *ura*3 mRNA directly by hybridization. Wild-type yeast and an isogenic *ura*1 mutant (derepressed 5-fold) were grown in excess uracil so that the growth rates were comparable. Total RNA was labeled by addition of [^{3}H]adenine to exponentially growing cells. After a short exposure (10–15 min) cells were either collected immediately or an excess of nonradioactive adenine was administered in order to observe mRNA decay. The chase became effective (i.e., [^{3}H]adenine no longer was incorporated into acid-insoluble RNA) within 10 min. Equilibrium labeling of RNA was done over at least three generations under conditions that leave at least 75% of the radioactivity unincorporated. Hybridization was carried out at least in triplicate and OMP decarboxylase specific activity was determined in the same cultures.

Table 2 shows the results of measurements of specific hybridization of RNA to fragment b (the *ura*3 gene) in the two strains, one wild type and one induced; specific enzyme activity in the same cultures is also given. There was a comparable difference (a factor of 4–5) in the level of hybridization when labeling of the RNA was either in a relatively short pulse (10–15 min) or over three generations (i.e., equilibrium). Other experiments (not shown) using different strains and different inducing conditions gave consistent results: the specific hybridization and the enzyme activity varied coordinately. Shorter pulse-labeling experiments (i.e., 2.5 min) also gave comparable differences, indicating that the differences in hybridization reflect differences in the synthesis of the RNA and not differences in decay rate.

In order to verify that the RNA that hybridizes to the clone 6 DNA behaves like a normal yeast mRNA, the apparent half-life was measured and compared with RNA hybridizing to two other hybrid plasmids that were found to be complementary to abundantly transcribed nonribosomal RNAs (6). As shown in Fig. 4, the *ura*3 mRNA had an apparent half-life of about 10.5 min, whereas the two putative "abundant messengers" had apparent half-lives of 7 and 14.5 min. All three mRNAs decayed exponentially with decay times within the range (3–30 min) found for different yeast mRNAs under various conditions (27–30).

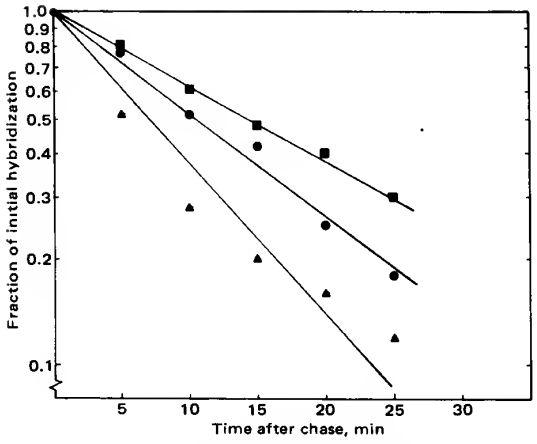


FIG. 4. Decay rates of yeast mRNA. RNA was pulse-labeled for 10 min with [^{3}H]adenine followed by a chase of unlabeled adenine. RNA was extracted from samples taken at intervals thereafter and hybridized to clone 6 (*ura*3), clone PL1-B11 ("abundant mRNA 1"), or clone PL13-H1 ("abundant mRNA2"). The latter two clones contain DNA complementary to unknown abundant unstable RNA and do not contain DNA complementary to ribosomal RNA (6). Hybridization is shown as a fraction of the total acid-insoluble radioactivity in the same volume of culture measured at each time point; this value increased less than 40% during the first 10 min of the chase and remained quite constant (increased slowly, less than 10%) thereafter. The range of values of cpm in *ura*3 mRNA [(cpm on clone 6 filter) − (cpm on pBR322 alone); mean (± SD) of three filters each] was 220 ± 18 to 10 ± 3 cpm. ■, Abundant mRNA 1, $t_{1/2}$ = 14.5 min; ●, *ura*3 mRNA, $t_{1/2}$ = 10.5 min; ▲, abundant mRNA 2, $t_{1/2}$ = 7 min.

DISCUSSION

The techniques of *in vitro* recombination have made it possible to isolate and purify eukaryotic structural genes. In this paper we have described the isolation and identification of an 1100-base-pair piece of yeast DNA that appears to contain the structural gene for yeast OMP decarboxylase, the *ura*3 gene. The gene was found by virtue of its expressibility in *E. coli*, a property previously described for various yeast genes encoding biosynthetic enzymes (2, 3).

The availability of structural gene DNA in abundance made possible hybridization experiments to test directly whether the

Table 2. OMP decarboxylase: mRNA and enzyme levels in induced and uninduced yeast strains

	Enzyme activity*	Specific hybridization under different conditions†				
		15-min PL			Equilibrium‡	
		Exp. 1	Exp. 2	2.5-min PL	Exp. 1	Exp. 2
ura$^+$	3.8	2.4	2.0	2.2§	0.22 ± 0.13	0.27 ± 0.14
*ura*1–21	19.4	10.7	9.2	9.4	0.88 ± 0.1	0.94 ± 0.1
Ratio	5.1	4.5	4.6	4.3	4.0	3.5

Cells were exponentially growing in excess (50 μg/ml) uracil in all cases.

* Shown as μmol/min per mg.

† PL, pulse label. Shown as fraction of total acid-insoluble ^{3}H-labeled RNA × 10^{-5}; in most cases, at least 10^6 cpm was applied to each filter.

‡ At three generations. These numbers are at the limit of significance; therefore, in each experiment four different filters were used for each sample. The filter-bound radioactivity was counted for 20 min and compared to that on four filters loaded only with vector (pB4322) DNA. Statistical analysis yielded accumulated errors (i.e., combined SD for blank and for sample) as indicated. Control experiments comparing vector blanks with filters lacking DNA showed no statistically significant difference. Other results in the table have computed statistical errors less than 10%.

§ A *ura* 2 mutant was used in this experiment; as shown previously (4), *ura* 2 mutants are not induced and have OMP decarboxylase levels identical to those of *ura*$^+$ strains.

regulation of the expression of the *ura*3 gene occurs at the level of transcription. We found that the level of *ura*3 mRNA as measured by hybridization to the structural gene DNA varies coordinately with the observed enzyme activity. It thus is possible to conclude that the *ura*3 gene is under transcriptional regulation.

Measurements of the degree of expression of the *ura*3 gene in *E. coli* have also been carried out. At the present time it seems clear that there is some variation depending upon the conditions of cell growth and upon the particular hybrid plasmid used. These variations could indicate that some kind of regulation might occur even in *E. coli*, but the nature of this variation is still not understood well enough to permit any firm conclusion.

Related to this observation of variability in expression of the *ura*3 gene in *E. coli* is the likelihood that the 1100-base-pair fragment of DNA contains a promoter recognized by *E. coli* RNA polymerase. At present there is no definitive proof of this and no indication as to whether the promoter used in *E. coli* bears any relationship to the normal promoter used by the relevant yeast RNA polymerase(s).

We also measured the half-life of the *ura*3 mRNA. The value we observed for this RNA (10.5 min) and for two abundantly transcribed RNAs (7 and 14.5 min) are in the range of published values for yeast mRNA decay rates (27–30). The variation among transcripts that we found is clearly significant, and the fact that these are direct measurements of particular chemical species should be emphasized. The fact that decay is exponential with little or no delay makes it somewhat less likely that the length of the poly(A) "tail" added to mRNA after transcription regulates its half-life. Nevertheless, such a hypothesis can now be tested because the two abundant messenger species measured in our experiments have such different decay rates.

The amount of *ura*3 mRNA we found in equilibrium labeling in the wild type is about 2×10^{-6} of total RNA. Given that the total messenger comprises about 1.5% of total RNA, the *ura*3 mRNA comprises about 1.3×10^{-4} of the total mRNA. If there are about 10^4 genes in yeast, the uninduced level of transcription of *ura*3 appears to be about average.

We thank Tom Petes for many essential materials and much helpful advice; Paula Grisafi, for superb technical assistance; Mark Rose, for help with some experiments; and Ron Davis and Gerry Fink, for access to unpublished information and methods. Work in Cambridge was supported by grants from the National Institutes of Health (GM18973 and GM21253) and from the American Cancer Society (VC-245). Work in Strasbourg was supported by Grant 77-1-073-1 from Institut National de la Santé et de la Recherche Médicale, by the Microbiology ATP No. 3070 from the Centre National de la Recherche Scientifique, and by a grant from the French Society for Medical Research.

1. Struhl, K., Cameron, J. R. & Davis, R. W. (1976) *Proc. Natl. Acad. Sci. USA* 73, 1471–1475.
2. Ratzkin, B. & Carbon, J. (1977) *Proc. Natl. Acad. Sci. USA* 74, 487–491.
3. Struhl, K. & Davis, R. W. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5255–5259.
4. Lacroute, F. (1968) *J. Bacteriol.* **95**, 824–832.
5. Bach, M. L. & Lacroute, F. (1972) *Mol. Gen. Genet.* **115**, 126–130.
6. Petes, T. D., Broach, J. R., Wensink, P. C., Hereford, L. M., Fink, G. R. & Botstein, D. (1978) *Gene*, in press.
7. NIH Guidelines for Recombinant DNA Research (1976) *Federal Regis.* **41**, (131), 27902–27943.
8. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H. & Falkow, S. (1977) in *DNA Insertion Elements, Plasmids and Episomes*, eds. Bukhari, A. I., Shapiro, J. A. & Adhya, S. L. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 686–687.
9. Miller, J. H. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
10. Mendel, M. & Higa, A. (1970) *J. Mol. Biol.* **53**, 159–162.
11. Sharp, P. A., Sugden, B. & Sambrook, J. (1973) *Biochemistry* **12**, 3055–3065.
12. Maniatis, T., Jeffrey, A. & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1184–1188.
13. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
14. Botchan, M., Topp, W. & Sambrook, J. (1976) *Cell* **9**, 269–287.
15. Petes, T. & Botstein, D. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5091–5095.
16. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 3240–3244.
17. Thomas, M., Cameron, J. & Davis, R. W. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4579–4583.
18. Wolcott, J. H. & Ross, C. (1966) *Biochim. Biophys. Acta* 122, 532–534.
19. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. (1951) *J. Biol. Chem.* **193**, 265–275.
20. Waldron, C. & Lacroute, F. (1975) *J. Bacteriol.* **122**, 855–865.
21. Baker, R. F. (1977) *Anal. Biochem.* 78, 569–571.
22. Gillespie, D. & Spiegelman, S. (1965) *J. Mol. Biol.* 12, 829–842.
23. Kourilsky, P., Mercereau, O., Gros, D. & Tremblay, G. (1974) *Biochimie* **56**, 1215–1221.
24. Birg, F., Favaloro, J. & Kamen, R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3138–3142.
25. Clewell, D. B. (1972) *J. Bacteriol.* **110**, 667–676.
26. Hinnen, A., Hicks, J. B. & Fink, G. R. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1929–1933.
27. Hutchison, H. T., Hartwell, L. H. & McLaughlin, C. S. (1969) *J. Bacteriol.* **99**, 807–814.
28. Tonnesen, T. & Friesen, J. D. (1973) *J. Bacteriol.* **115**, 889–896.
29. Cannon, M., Davies, J. E. & Jimenez, A. (1973) *FEBS Lett.* 32, 277–280.
30. Lawther, R. P., Phillips, S. L. & Cooper, T. G. (1975) *Mol. Gen. Genet.* **137** (2), 89–99.

TAB S

Nature Vol. 279 3 May 1979

Expression in *Escherichia coli* of hepatitis B virus DNA sequences cloned in plasmid pBR322

C. J. Burrell*, Patricia Mackay
Department of Bacteriology, University of Edinburgh Medical School, Edinburgh UK

P. J. Greenaway, P. H. Hofschneider†
Microbiological Research Establishment, Porton Down, Salisbury, Wiltshire, UK

K. Murray
Department of Molecular Biology, University of Edinburgh, King's Buildings, Edinburgh UK

Fragments of hepatitis B virus DNA isolated from Dane particles have been inserted into the Escherichia coli *plasmid pBR322 and cloned. Cells carrying the hybrid plasmid synthesise antigenic material that reacts specifically with antisera to hepatitis B viral antigens.*

INFECTION with hepatitis B virus (HBV) is widespread in man. Between 3 and 15% of healthy blood donors in Western Europe and the US show serological evidence of past infection and about 0.1% are chronic carriers of the virus. In many African and Asian countries the prevalence is much higher and the majority of the adult population have been infected, while 5–10% of the population are chronically infected[1]. Most infections are subclinical and are followed by apparently complete recovery with the development of virus-specific antibody. However, a significant proportion of infections (probably 1–5%) may produce chronic sequelae including persistent infection, chronic hepatitis of various types, cirrhosis and possibly primary liver cancer.

Plasma from some blood donors and patients infected with HBV contains 42-nm spherical particles (Dane particles)[2] which have serological and biochemical properties, suggesting that they are the infective virions of hepatitis B. These have an outer envelope containing the hepatitis B surface antigen (HBsAg) and an inner core (diameter 27 nm) bearing a second unrelated antigen, the hepatitis B core antigen (HBcAg). Within the core is a double-stranded circular DNA molecule of molecular weight $\sim 2 \times 10^6$ which has a large variable gap in one strand, and an endogenous DNA-dependent DNA polymerase activity that can fill in this gap in *in vitro* reactions[3]. There is some evidence that the total virus genome length may be around one-third greater than the 2×10^6 daltons found in single molecules in which case productive infection of a cell may require simultaneous infection by at least two genetically different particles[3]. A third antigen, the hepatitis B e antigen (HBeAg), which is probably also virus-coded, is found free in the plasma of some infected individuals and possibly also in association with Dane particles. Passively or actively acquired antibody to HBsAg (anti-HBs) confers some immunity to subsequent HBV challenge, and prototype vaccines composed of purified inactivated HBsAg prepared from the plasma of infected carriers are being evaluated[4]. However, the virus cannot be grown in tissue culture and normally infects only man and apes. This means that molecular studies of the virus and its genome have been based on the limited amounts of material obtainable from the plasma of infected individuals. Such studies could be advanced considerably by insertion of HBV DNA into a bacterial plasmid or phage to allow its production in quantity from cloned, purified single molecules. Such clones would also be useful for studies of the expression of HBV gene products in bacterial cells, and for large-scale production of HBV DNA and viral antigens for diagnostic purposes and, possibly, vaccine production.

Restriction of HBV DNA

The published information on digestion products of HBV DNA (Dane particle DNA) with restriction endonucleases[5,6] is limited and so additional analyses of this type were carried out before attempting cloning experiments. The amount of HBV DNA available was limited (about 90 ng from 5 ml of plasma), so the DNA was first labelled with ^{32}P in the endogenous DNA polymerase reaction[3]. DNA from bacteriophage λ was then added as a carrier to titrate the restriction enzyme and to provide reference fragments of known size. The digests were analysed by electrophoresis in agarose gels and the results of some of these experiments are shown in Fig. 1.

The heterogeneity of the undigested labelled HBV DNA (tracks *b* and *e*) precluded a detailed analysis of the restriction digests. The multiple bands of rather similar size could be due in part to the presence of linear and circular forms of otherwise similar molecules, in part to differing degrees of repair synthesis, although repeated DNA preparations from the same plasma sample gave reproducible patterns; the heterogeneity could also represent a true molecular dispersity from a mixed population of virions. HBV DNA preparations from several individual donors were all heterogeneous and differed slightly from each other (an example is included in Fig. 1; tracks *a*, *b* and *c* compared with *d*, *e* and *f*) both before and after digestion with various restriction enzymes. Heterogeneity has also been observed by Landers *et al.*[6] in the products of the endogenous polymerase reaction which comprised two principal radioactive components representing linear and circular molecules with additional minor components due to incomplete repair. Examination of the HBV DNA (Fig. 1, tracks *b* and *e*) by electron microscopy showed that the population contained circular molecules of MW $\sim 2 \times 10^6$ and linear molecules ranging from 0.5 to 10×10^6. The linear molecules represented about three times the concentration of circular molecules and only a few of them had a MW around 2×10^6, the majority being between 0.5 and 1×10^6. The heterogeneity observed on electrophoresis was, therefore, not due to circular and linear forms of similar length and few, if any, of the linear molecules were labelled in the polymerase reactions.

* Present address: Division of Virology, I.M.V.S., Box 14, Rundle St. Post Office, Adelaide, S. Australia.
† Permanent address: Max Planck Institute for Biochemistry, Am Klopferspitz, Martinsried bei Munich, FRG.

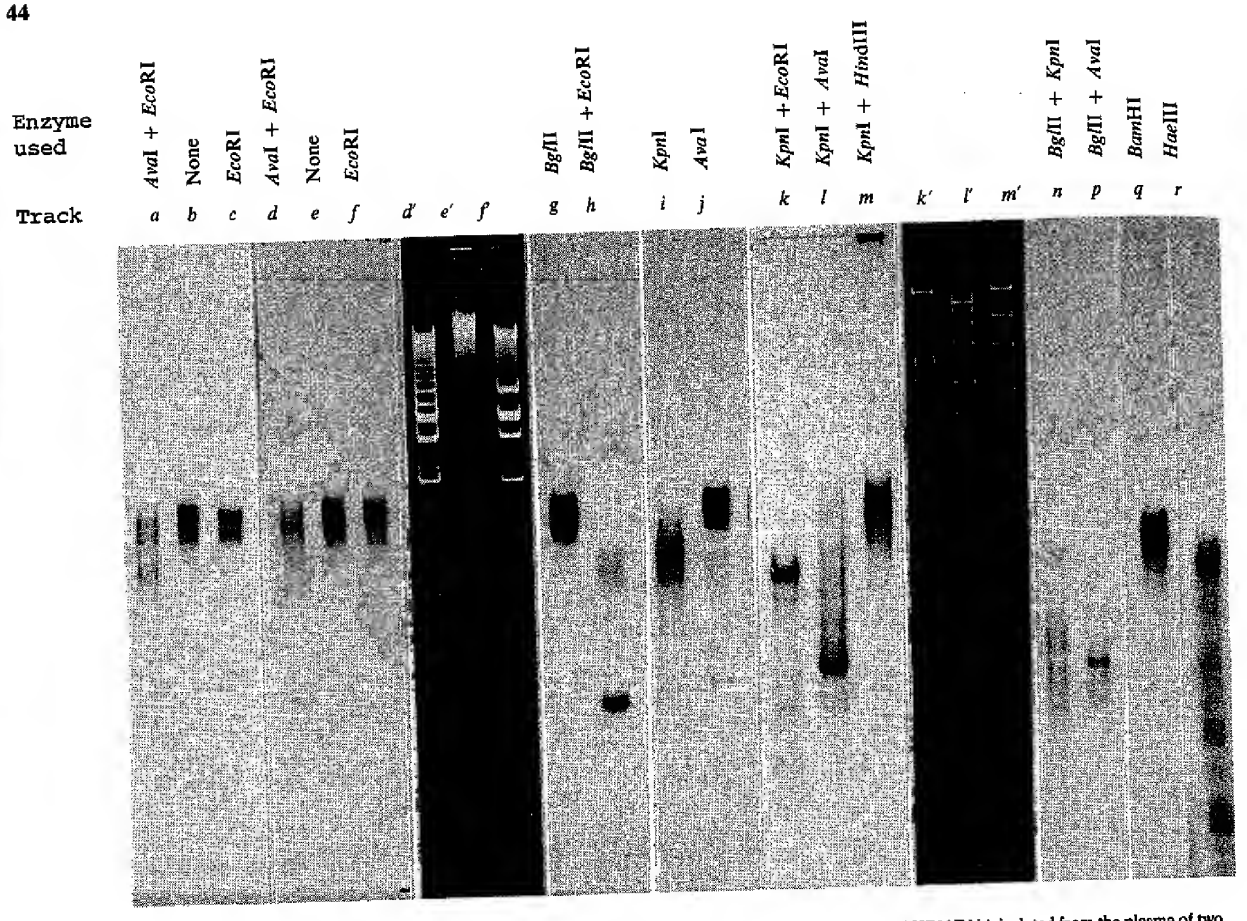


Fig. 1 Autoradiographs of restriction enzyme digests of HBV DNA after electrophoresis in agarose gels. Examples of HBV DNA isolated from the plasma of two blood donors are shown, the samples in tracks *a*, *b* and *c* from one donor being directly comparable with those in tracks *d*, *e* and *f* from the other; the sample used for the digests in tracks *d*, *e* and *f* was also used for all the other digests shown, and also for the preparative experiments for cloning. In addition to the autoradiographs, the figure includes examples, indicated by primed letters, of the gels stained with ethidium bromide to reveal fragments of the λ^+ DNA included with the HBV DNA in the restriction reactions. HBV DNA was prepared from Dane particles isolated by two cycles of ultracentrifugation from clarified plasma obtained from individual HBsAg-positive blood donors essentially as described by Landers *et al.*[6]. The single-strand gaps in the DNA molecules were repaired and the DNA radioactively labelled in reactions with the endogenous DNA polymerase in which ^{3}H-dCTP and ^{3}H-dTTP (22 and 30 mCi μmol^{-1}, respectively; Radiochemical Centre Amersham) or ^{32}P-dGTP (300 mCi μmol^{-1}) were included[6]. The released core particles were further purified by sedimentation through 30% w/v sucrose solution and DNA was isolated by treatment with proteinase K (Boehringer-Mannheim; 2 mg ml^{-1} in 0.6% SDS) followed by phenol extraction and dialysis. Restriction enzyme digests (37 °C. 1.5 h) were carried out in 10 mM Tris-HCl, *p*H 7.5, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 40 mM NaCl after addition of phage λ DNA (0.5–1 μg per reaction) and the reactions were stopped by heating at 70 °C for 5 min. The samples were then applied to 1% w/v agarose gels[23] in 0.04 M Tris-acetate, *p*H 8.2 for electrophoresis (35 mA, 8 h). Gels were stained with ethidium bromide and photographed under UV light[24] and then either placed in alkali to denature DNA fragments for transfer to cellulose nitrate membrane filters[13] or dried on Whatman 3MM paper for autoradiography. The MW of the fragments of HBV DNA was estimated from their electrophoretic mobility[24] with reference to fragments of λ^+ DNA in digests with R.*Eco*RI and R.*Hin*dIII included in the same gel[25]. These results are included in Fig. 2 which gives a provisional map of some of the restriction targets relative to each other.

Digestion of HBV DNA with R*Eco*RI or R.*Bgl*II changed the pattern of the major bands only slightly (Fig. 1 tracks *c*, *f* and *g*) whereas digestion of the DNA with these two enzymes together (track *h*) gave a radioactive fragment, MW 0.75×10^6 as the principal component, with little of the original DNA remaining. This is consistent with the introduction of a single break in circular molecules by either enzyme alone to give intact linear molecules; when the two enzymes acted together two fragments were formed, the smaller containing the entire labelled region. On digestion with R.*Kpn*I (track *i*) almost the whole group of bands was displaced down the gel corresponding with the loss from each of a fragment of MW $\sim0.4\times10^6$ but no corresponding radioactive fragment was found. This suggests that HBV DNA may contain at least two targets for R.*Kpn*I located outside the region repaired in the reaction with DNA polymerase and spanning a sequence common to all the molecules. The alternative explanation requiring cleavage of circular molecules at a single target to give linear molecules is less likely in view of the behaviour observed in the R.*Eco*RI and R.*Bgl*II digests. R.*Hae*III furnished a spectrum of fragments with a range of sizes (track *r*) in a pattern broadly similar to that published[5,6]. Digestion of HBV DNA with R.*Ava*I and with R.*Bam*HI gave several radioactive fragments (tracks *j* and *g*); a major product of the R.*Ava*I digest had a MW of 0.88×10^6, while radioactive fragments of 1.2×10^6 and 1.8×10^6 occurred in the R.*Bam*HI digests; in other digests with R.*Bam*HI smaller fragments were also observed. These results, together with the principal radioactive products found in various digests with pairs of restriction enzymes (for example, R.*Eco*RI and R.*Ava*I, track *a* or *d*) are summarised in Fig. 2. Within the constraints imposed by the dispersity of the HBV DNA preparations used, they provided the approximation of relative positions of restriction targets as shown.

The results imply that, in most molecules, the single-stranded gap repaired by the endogenous DNA polymerase lies within a relatively constant region of the DNA sequence; Landers *et al.*[6] similarly concluded, from an analysis of R.*Hae*III digests after different polymerase reaction times, that DNA repair took place largely within the same one-third to one-half of the total DNA sequence, although initiation of DNA repair could occur at variable sites within this region.

Digests of HBV DNA with R.*Eco*RI or R.*Bgl*II offer the possibility of cloning the entire HBV genome, while major fragments might be cloned from digests with R.*Bam*HI, R.*Kpn*I

or R.*Ava*I, or from various double digests. For structural studies of the HBV genome it is desirable to clone the entire DNA molecule, but clones covering a range of fragments could well be more useful for attempts to demonstrate the expression of HBV sequences in *E. coli*. Micromethods based on radioimmunoassay can be applied to individual bacterial colonies or phage plaques for the detection of specific polypeptides[7]. As many eukaryotic genes will not be expressed in prokaryotic cells it appeared desirable to insert HBV DNA fragments within a prokaryotic gene so as to produce a fused polypeptide; for this purpose the R.*Pst* target in the *E. coli* plasmid pBR322 has been used successfully[8].

Cloning of HBV DNA fragments in pBR322

HBV DNA isolated from Dane particles from a single HBsAg positive, HBeAg positive donor (serotype *adyw*) was labelled to a low specific radioactivity with ^{3}H in a repair reaction with the endogenous polymerase in order to facilitate its handling. This preparation was then variously digested with R.*Eco*RI, R.*Bam*HI, R.*Bgl*II, R.*Kpn*I and R.*Ava*I. Portions of the R.*Eco*RI and R.*Bam*HI digests were used for insertion at the respective sites of appropriately restricted pBR322 DNA by annealing and ligation with T4 DNA ligase. The fragments from the other restriction enzyme digests and the remainder of the R.*Eco*RI and R.*Bam* digests were treated with polynucleotide terminal transferase for addition of 3′ oligo (dC) sequences[9]. These fragments were annealed to pBR322 DNA to which oligo (dG) sequences had been attached after cleavage with R.*Pst*. The DNA preparations were then used to transform competent cultures of *E. coli* HB101 and transformants were screened for the acquisition of HBV sequences on the following basis. Cells transformed with recombinants made using the R.*Pst* site in pBR322 retained their resistance to tetracycline, but became sensitive to ampicillin. Cells transformed with recombinants made using the R.*Bam*HI site in pBR322 became sensitive to tetracycline, but retained their resistance to ampicillin, while cells transformed with DNA cloned using the R.*Eco*RI site remained resistant to both antibiotics[10] and were detected by colony hybridisation[11] with ^{32}P-labelled DNA from Dane particles. The presence of HBV DNA sequences in colonies where antibiotic resistance and sensitivities indicated the insertion of additional DNA into the plasmid was confirmed by colony hybridisation.

Characterisation of pBR322–HBV hybrids

Plasmid DNA isolated from cell cultures that had been treated with chloramphenicol to amplify plasmid production[12] was analysed by gel electrophoresis before and after digestion with restriction endonucleases.

DNA fragments from several of the gels were transferred to cellulose nitrate filters for hybridisation[13] with HBV DNA labelled with ^{32}P by the endogenous polymerase reaction. Examples of these results are shown in Fig. 3 and some of the characteristics of the cloned segments are given in Table 1. The hybridisation observed with appropriate fragments does not establish conclusively that the cloned sequences were HBV-specific, for the cloned DNA and ^{32}P-labelled probe had been prepared from the same plasma sample. Thus contaminating non-viral DNA fragments that had been cloned inadvertently would also have been detected if the same non-viral DNA sequence was labelled by the endogenous polymerase; however, in other experiments such ^{32}P-labelled probes hybridised with DNA from HBV-infected liver tisue, but not with DNA from normal human liver. Furthermore, preparations of ^{32}P-labelled DNA made from Dane particles purified from four different blood donors gave the same patterns when hybridised against restricted DNA fragments from the recombinant plasmids, thus strengthening the view that the cloned sequences were in fact HBV DNA.

The results presented in Fig. 3 are examples taken from the large number of recombinant colonies obtained from the transformation experiments and relate to colonies described below.

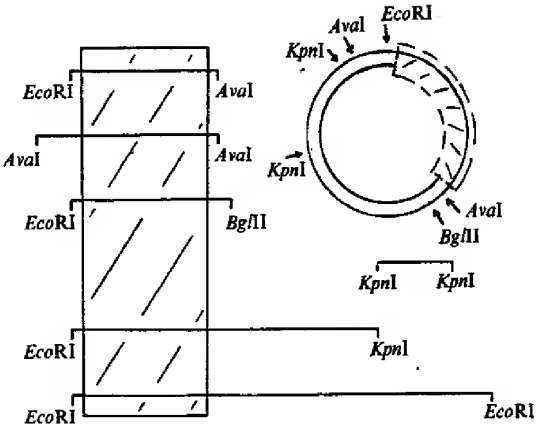


Fig. 2 The size and approximate relative order in the HBV genome of some of the major fragments in various restriction enzyme digests (Fig. 1) of HBV DNA. The site for R.*Eco*RI is taken arbitrarily as a reference point for the circular map. The shaded area indicates the region labelled with ^{32}P in the endogenous repair reaction.

Expression of HBV sequences

A population of the colonies carrying putative recombinant plasmids, selected on the basis of drug resistance and sensitivity characteristics, was screened for the production of HBV antigens. Three test systems were used in the solid phase method of radioimmunoassay described by Broome and Gilbert[7] which uses polyvinyl disks coated with IgG from specific antisera. These disks are placed in contact with cells producing antigen which binds to the IgG surface. When subsequently incubated with ^{125}I-labelled IgG from the same antiserum, the bound antigen retains the label and can be detected readily by autoradiography. The sera used were anti-HBs (human and hyperimmune animal sera), anti-HBc (human sera containing HBsAg and HBeAg and selected on the basis of a high anti-HBc titre) and anti-HBc and anti-HBe together (human sera containing HBsAg and high levels of anti-HBc and anti-HBe).

Clear positive results were obtained in the two test systems containing anti-HBc. Of some 350 colonies tested, 13 gave intense spots on autoradiography with the anti-HBc+anti-HBe antibodies (Fig. 4*a*). Most of these colonies remained strongly positive when re-tested in both the anti-HBc and anti-HBc+anti-HBe assay systems (Fig. 4*b*). None of the clones giving a negative response in the initial screening was subsequently positive, but some clones that were positive initially gave a negative result after subculturing, which probably reflects instability of some of the hybrid plasmids. All of the clones that were positive in the anti-HBc+anti-HBe system were positive when tested with anti-HBc alone, implying that none was producing detectable levels of HBeAg. If lysis of the bacterial colonies with phage λ before radioimmunoassay was omitted, only a very faint outline of the positive colonies was discernible (Fig. 4*c*). This is consistent with the presence of the antigen as a periplasmic polypeptide fused to the major part of β-lactamase (penicillinase) as anticipated. The sensitivity of this assay was shown to be comparable with that attainable with the same reagents in a solid phase microtitre well assay[14] by titration of 10 μl samples (diluted progressively from 1 in 400) of semi-purified HBcAg from human liver[15].

The immunological specificity of the reactions was confirmed by the following observations. Replicate assays with different anti-HBc sera identified the same positively reacting clones. The positive reactions in the HBcAg assay were abolished if the polyvinyl disks were coated with normal human IgG instead of specific anti-HBc IgG, or if the normal human serum used as diluent for the ^{125}I-labelled antibody was replaced by anti-HBc-positive serum from a different donor (by competition with excess unlabelled anti-HBc), or if ^{125}I-anti-HBS replaced ^{125}I-anti-HBc in the assay with the anti-HBc coated disks. Finally,

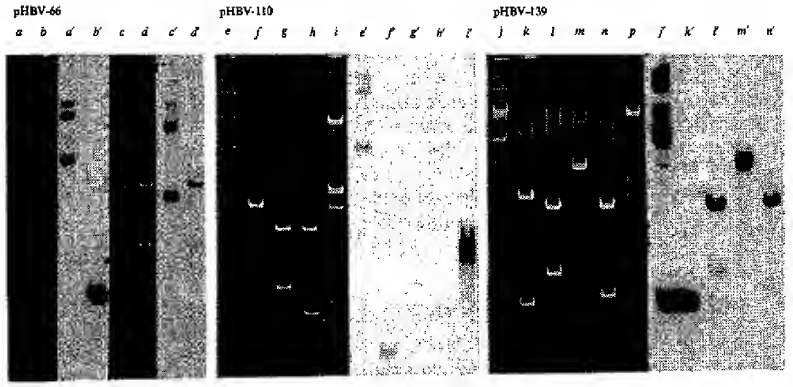


Fig. 3 Electrophoretic separation in agarose gels of restriction enzyme digests of recombinant plasmids comprising pBR322 and HBV DNA sequences. The examples shown are of plasmids (Table 1) that elicit production of antigenic material that reacts specifically with antibodies to HBcAg (Fig. 4). The left-hand panels show DNA fragments revealed by staining with ethidium bromide and photography under UV light. After photography, gels were soaked in alkali for denaturation of the DNA fragments which were then transferred to cellulose nitrate membrane filters[13] for hybridisation with ^{32}P-labelled HBV DNA and autoradiography; the right hand panels show the corresponding radioautographs. HBV DNA (~0.5 μg) labelled with ^{3}H was prepared from 40 ml clarified blood plasma from a donor with a high titre of HBsAg, serotype *adyw* as described in the legend to Fig. 1 and ref. 6. In one experiment 0.5 μg *E. coli* DNA was added as a carrier, but in a second the carrier was omitted. The DNA was divided into six portions for various restriction enzyme digests. Portions of R.*Eco*RI and R.*Bam*HI digests were incubated with pBR322 digested with the same enzyme in reactions with T4DNA ligase[26] (1 U ml^{-1}, 10 °C for 3 h followed by storage at 0 °C) in 66 mM Tris-HCl *p*H 7.2, 10 mM $MgCl_2$, 40 mM NaCl, 0.2 mM EDTA, 0.1 mM ATP, 10 mM 2-mercaptoethanol. The remainder of the R.*Eco*RI and R.*Bam*HI digests as well as R.*Kpn*I and R.*Bgl*II digests were used in terminal transferase reactions, but with the exception of the R.*Kpn*I digest the samples were first incubated with phage λ exonuclease (1.5 h at 0 °C in 50 mM Na glycinate, *p*H 9.5, 5 mM $MgCl_2$, 50 μg ml^{-1} bovine serum albumin (BSA) followed by phenol extraction and recovery of the DNA by ethanol precipitation) to remove the 5′ single-stranded projections left by the restriction enzymes. Poly(dC)sequences were attached to the 3′ termini by incubation with polynucleotide terminal transferase[9] (250 U ml^{-1} for 10–20 min at 27 °C) in 15 μl 100 mM potassium cacodylate, *p*H 7.0, 1 mM $CoCl_2$, 1 mM dCTP, 50 μg ml^{-1} BSA, and the reactions stopped by addition of excess EDTA. An approximately molar equivalent of pBR322 DNA digested with R.*Pst* and incubated in similar reactions with dGTP instead of dCTP (from Dr J. van den Berg) was then added and the samples diluted to 50 μl for annealing in 50 mM Tris-HCl, *p*H 7.5, 100 mM NaCl, 5 mM EDTA, heated to 60 °C and gradually cooled to room temperature over a period of about 5 h. Aliquots of the solutions (10 μl) were then incubated with competent cultures (0.1 μl) of *E. coli* HB101 prepared as described by Lederberg and Cohen[27] and incubated overnight at 37 °C on L-agar plates containing tetracycline (20 μg ml^{-1}) or ampicillin (50 μg ml^{-1}). Colonies were subcultured on to Millipore filters supported on appropriate agar plates to test their resistance or sensitivity to the two antibiotics and for colony hybridisation[11] with ^{32}P-labelled HBV DNA from Dane particles. Colonies giving positive hybridisation reactions were grown in liquid culture and then shaken overnight at 37 °C after addition of chloramphenicol (170 μg ml^{-1}) to amplify the number of plasmids within the cells which were then collected by centrifugation and treated with lysozyme and EDTA[12]. The resulting spheroplasts were lysed with Triton X-100 and the plasmid recovered by equilibrium centrifugation in CsCl solution (0.95 g per ml lysate) containing ethidium bromide (200 μg ml^{-1}). The plasmid bands were collected, extracted with propan-1-ol saturated with aqueous CsCl, dialysed, and samples digested with restriction endonucleases as described in the legend to Fig. 1. The results shown are from the following hybrid plasmids: pHBV-66 in tracks *a*, *b*, *c* and *d* in which *a* is the undigested plasmid and *b*, *c* and *d* are digests with R.*Pst*, R.*Kpn*I and R.*Bam*HI, respectively; pHBV-110 in tracks *e*, *f*, *g* and *h*, in which *e* is the undigested plasmid and *f*, *g* and *h* are digests with R.*Pst*, R.*Bam*HI and R.*Bam*HI + R.*Eco*RI, respectively; pHBV-139 in tracks *j*, *k*, *l*, *m* and *n*, in which *j* is the undigested plasmid and *k*, *l*, *m* and *n* are digests with R.*Pst*, R.*Bam*I, R.*Eco*RI, and R.*Bam*HI + R.*Eco*RI, respectively. Tracks *i* and *p* show reference digests of λ^{+} DNA with R.*Eco*RI + R.*Hin*dIII[25] which also contained ^{32}P-labelled undigested DNA from Dane particles. The primed letters identify the samples on the autoradiograph, made on Kodak X-omat H X-ray film with an intensifying screen, of the cellulose nitrate filters after hybridisation with HBV DNA.

absorption of the radioactive anti-HBc with semipurified HBcAg[15], by overnight incubation at 4 °C followed by centrifugation to remove the excess antigen, abolished the positive result in the radioimmunoassay. Thus the detection of an as yet unidentified HBV-specific antigen by interaction with its specific antibody present in the sera of HBV-infected individuals remains a formal possibility, but it is unlikely, and the results are wholly consistent with the observed activity being that of HBcAg. The HBcAg polypeptide detected in the colonies, however, is unlikely to be identical with that occurring naturally for the cloning experiment was such as to produce polypeptides linked to β-lactamase[8]. However, it has not been established that this is the case and it remains possible that translation of β-lactamase sequences could be terminated and followed by reinitiation, but against this is the occurrence of clones that do not exhibit HBcAg activity, but which contain HBV DNA fragments similar to those that do. The point will be clarified by DNA sequence determination, which is in progress. All of the 13 HBcAg-positive clones had been made using the R.*Pst* site in pBR322, five with fragments from R.*Bam*HI digests of HBV DNA and eight from R.*Kpn*I digests. Not all of the plasmids have been isolated for analysis, but the three illustrated in Fig. 3 were all from cells giving positive reactions for HBcAg. The smallest fragment of HBV DNA in these plasmids (Table 1) was about 0.95×10^6 daltons, but others with an HBV fragment about half this size also gave positive reactions for HBcAg; values reported[3] for the MW of the naturally occurring HBcAg range from 17,000 to 80,000.

In an equivalent assay for HBsAg, similar, but faint positive reactions were obtained with four clones which are being analysed further. One might expect detection of expression of serological activity to be more difficult with HBsAg than with HBcAg as its protein moiety is markedly hydrophobic and hence

Table 1 Some properties of the plasmids shown in Fig. 3

Hybrid plasmid	MW of fragment excised by R.*Pst* ×10^{-6}	Targets for restriction enzymes within the HBV sequences R.*Eco*RI	R.*Bam*HI	R.*Ava*I
pHBV-66	1.2	–	+	+
pHBV-110	0.95	–	+	+
pHBV-139	1.16	–	+	+

In all the plasmids, the site for R.*Bam*HI within the HBV sequence is located about 0.7×10^6 daltons from the R.*Pst* site near the R.*Eco*RI site.

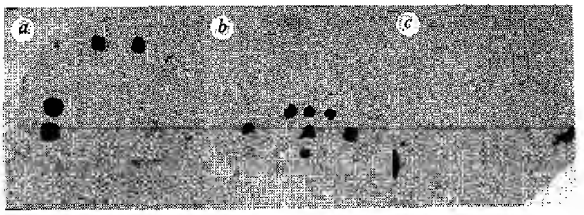


Fig. 4 Autoradiographs showing the detection by radioimmunoassay of bacterial colonies expressing HBcAg. *a*, Four HBcAg-positive colonies amongst 52 colonies examined on one plate. *b*, The result obtained when initially positively reacting colonies, together with a random selection of colonies giving a negative reaction, were subcultured from stock plates and re-tested. *c*, A duplicate of *b* in which lysis of the colonies with phage λ before radioimmunoassay was omitted. Colonies of the bacteria were grown at 37 °C overnight on Millipore filters supported on nutrient plates containing tetracycline (to maintain selection for the plasmid). The cells were lysed by imprinting the filters for a few minutes on a lawn of bacteria confluently lysed with a virulent derivative of phage λ and then incubating further for several hours at 37 °C. Polyvinyl disks coated with anti-HBc + anti-HBe specific human IgG (60 μg ml^{-1}) were placed face down on the colonies, which had obviously lysed, and incubated at 4 °C for 3–4 h. The disks were then removed and washed thoroughly and vigorously to remove the considerable quantity of adherent viscous bacterial debris. Finally the disks were incubated overnight at 4 °C with homologous ^{125}I-labelled IgG (10^5 c.p.s. per μg, 2×10^4 c.p.s. per ml), washed thoroughly and autoradiographed on Kodak Blue X-ray film exposed for 2 d with an intensifying screen.

tends to remain associated with lipid[16], and some experiments (but not others) suggest that carbohydrate residues on HBsAg glycoprotein may be required for full serological activity[17,18].

Conclusions and further implications

A large number of hybrid DNA molecules comprising pBR322 and various fragments of the HBV genome have been cloned and propagated in *E. coli*. It thus becomes possible to produce HBV DNA in the quantities required for detailed structural and sequence analysis and location of the various coding sequences in the viral genome. Such analysis with DNA from several independent isolates and clones will explain the basis of the heterogeneity found in DNA from Dane particles. The DNA will also be useful for further genetic manipulation related to studies of expression and as a source of a highly radioactively labelled probe for hybridisation experiments for both diagnostic purposes and fundamental studies.

When inserted within a normal coding sequence of the *E. coli* plasmid the DNA from hepatitis B virus, which normally infects only man and apes, can be expressed to give serologically active translation products. This suggests that at least the region of the HBV genome coding for the amino acid sequence necessary for serological activity probably does not contain inserted sequences, or 'introns', which have now been found in a number of eukaryotic and viral genes[19–22], for it is widely believed, although not proved, that *E. coli* cannot process transcripts of these sequences. None of the hybrid plasmids so far examined contains a target for *R.Eco*RI within the HBV sequences. It was thus possible to insert the entire plasmid into a derivative of bacteriophage λ using this restriction site and plaques of the recombinant phage gave positive reactions in the disk radioimmunoassay (results not shown). The HBV sequences may well prove more stable when propagated within the phage genome, especially as a lysogen, and appropriate exploitation of the phage regulatory systems should permit significantly increased yields of the antigens from *E. coli*, raising the possibility of large scale antigen production for diagnostic purposes and development of vaccines. The insertion of HBV DNA into a phage λ derivative has been described very recently by Fritsch *et al.*[28].

We thank Professor B. P. Marmion, Dr R. J. C. Harris and many of the staff at MRE, Porton, for their help, Glynis Leadbetter for radioimmunoassay reagents, Archie Barr and Anthony Keen for HBsAg-positive plasma, Johan van den Berg and Professor Charles Weissmann for reagents and help with terminal transferase reactions, Eric Gowans and Pamela Beattie for electron microscopy, Carol White and Eva Zehelein for technical assistance. This work was assisted in part by a grant from the Scottish Home and Health Department and was supported partly by Biogen S.A. Cloning experiments were performed in CIV conditions as recommended by GMAG; work on DNA from Dane particles was done in a CI laboratory and DNA from cells transformed with hybrid plasmids in a CII laboratory.

Received 22 February; accepted 16 March 1979.

1. Szmuness, W., Harley, E. J., Ikram, H. & Stevens, C. E. in *Viral Hepatitis* (eds Vyas, G. N., Cohen, S. N. and Schmid, R.) (Franklin Institute Press, Philadelphia, 1978).
2. Dane, D. S., Cameron, C. H. & Briggs, M. *Lancet* **i**, 695–698 (1970).
3. Robinson, W. S. *A. Rev. Microbiol.* **31**, 357–377 (1977).
4. Purcell, R. H. & Gerin, J. L. *Am. J. clin. Path.* **70**, 159–169 (1978).
5. Summers, J., O'Connell, A. & Millman, I. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4597–4601 (1975).
6. Landers, T. A., Greenberg, H. B. & Robinson, W. S. *J. Virol.* **23**, 368–376 (1977).
7. Broome, S. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2746–2749 (1978).
8. Villa-Komaroff, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3727–3731 (1978).
9. Roychoudhury, R., Jay, E. & Wu, R. *Nucleic Acids Res.* **3**, 101–116 (1976).
10. Bolivar, F. *et al. Gene*, **2**, 95–133 (1977).
11. Grunstein, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961–3965 (1975).
12. Clewell, D. B. *J. Bact.* **110**, 667–676 (1972).
13. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
14. Purcell, R. H., Wong, D. C., Alter, H. J. & Holland, P. V. *Appl. Microbiol.* **26**, 478–484 (1973).
15. Cohen, B. J. & Cossart, Y. E. *J. clin. Path.* **30**, 709–713 (1977).
16. Howard, C. R. & Burrell, C. J. *Progr. med. Virol.* **22**, 36–103 (1976).
17. Burrell, C. J., Proudfoot, E., Keen, G. A. & Marmion, B. P. *Nature new Biol.* **243**, 260–262 (1973).
18. Shiraishi, H., Shirachi, R., Ishida, N. & Sekini, T. *J. gen. Virol.* **38**, 363–368 (1978).
19. Jeffreys, A. J. & Flavell, R. A. *Cell* **12**, 1097–1108 (1977).
20. Tilghman, S. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 725–729 (1978).
21. Breathnach, R., Mandel, J. L. & Chambon, P. *Nature* **270**, 314–319 (1978).
22. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. *Cell* **12**, 1–8 (1977).
23. Hayward, G. S. & Smith, M. G. *J. molec. Biol.* **63**, 383–395 (1972).
24. Sharp, P. A., Sugden, W. & Sambrook, J. *Biochemistry* **12**, 3055–3063 (1973).
25. Murray, K. & Murray, N. E. *J. molec. Biol.* **98**, 551–564 (1975).
26. Weiss, B. *Meth. Enzym.* XXI D, 319 (1971).
27. Lederberg, E. M. & Cohen, S. N. *J. Bact.* **119**, 1072–1074 (1974).
28. Fritsch, A., Pourcel, C., Charnay, P. & Tiollais, P., *C. r. hébd. Acad. Sci. Paris* **287**, 1453–1456 (1978).

letters

A 3-s delay in an optical burst from X-ray burst source MXB1735−44

THE discovery of optical bursts from the X-ray burst source MXB1735−44 was reported in a previous article[1]. One burst event was detected in both X rays and optical light. We showed (in Fig. 1 of the earlier article) that the optical burst is delayed by ≈1.5 s relative to the X ray burst. We concluded that to within the accuracy of the SAS 3 quick-look data timing (±1 s), the onset of the optical burst was coincident with that in X rays. A few months after the completion of our earlier paper, SAS 3 production data became available. We here report a re-analysis of the arrival times of the X ray and optical bursts based on these data which contain the universal time to an accuracy of ≲5 ms. We now find (as shown in Fig. 1) that the optical burst is delayed significantly by 2.8 s.

The optical observations were performed using the 1.5-m telescope at the Cerro Tololo Interamerican Observatory and a red-blocked RCA 31034 photomultiplier tube which had an effective bandpass including the transmission of the atmosphere of 3,100–5,500 Å (FWHM). The absolute time, which was maintained to within 10 ms of UT by monitoring WWV, and the optical data in 100-ms integrations were recorded on a digital tape. Simultaneous X-ray observations were made with the horizontal tube detectors aboard the SAS 3 observatory[2]. The three energy channels of interest are 3–6 keV, 6–12 keV and 8–19 keV which have time resolutions of 0.42 s, 0.83 s and 0.83 s, respectively. The X-ray production data are assigned absolute times in a three-step process. First, real time X-ray data and timing data from the spacecraft clock are tagged with the absolute time at the NASA ground station in Quito, Ecuador, by recording them simultaneously with the output of a UT caesium clock. Second, in the production analysis at Goddard Space Flight Center the real time data are used to assign UT to the bulk of the data which are played back from the on-board tape recorder. Third, corrections for the variable VHF propagation time from the satellite to the ground station are made using a solution on four days of radar-Doppler track-

TAB T

ent with the binaural estimates for the same stimulus conditions.

Although our measurements represent only a limited investigation of higher-frequency combination tones, we feel confident in concluding that, at least for low primary frequencies, such combination tones exist 20 to 40 dB below the primary levels. As others have suggested (*2*, *8*), the difficulty in identifying their presence appears to be largely due to the upward masking by the primary tones. Our observations indicate that the relative level of higher-frequency combination tones can be quite high. Thus, even though they are not clearly audible, they may be of some importance in identifying the nonlinear mechanism from which they arise. In this respect, the presence of higher-frequency combination tones brings the psychoacoustic results into closer agreement with physiological investigations of cochlear microphonic in which distortion components above and below the primary frequencies are routinely measured (*14*).

P. M. ZUREK
R. M. SACHS

Central Institute for the Deaf,
818 South Euclid,
St. Louis, Missouri 63110

References and Notes

1. H. L. F. von Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. A. J. Ellis, Transl. (from ed. 4, 1877) (Dover, New York, 1954) p. 158.
2. J. L. Goldstein, *J. Acoust. Soc. Am.* **41**, 676 (1967).
3. J. L. Hall, *ibid.* **51**, 1863 (1972); G. F. Smoorenburg, *ibid.* **52**, 615 (1972).
4. R. Plomp, *ibid.* **37**, 1110 (1965).
5. M. F. Meyer, *How We Hear* (Branford, Boston, 1950), chap. 15.
6. J. Erdreich and T. D. Clack, *J. Acoust. Soc. Am.* **55** (Suppl.), 7 (1974).
7. J. Erdreich, *ibid.* **63**, 1252 (1977).
8. J. W. Strutt (Lord Rayleigh), *The Theory of Sound* (Dover, New York, 1945) vol. 2, p. 460.
9. E. Zwicker, *Acustica* **5**, 67 (1955).
10. M. R. Schroeder, *J. Acoust. Soc. Am.* **46**, 1488 (1969); J. L. Hall, *ibid.* **56**, 1818 (1974); M. R. Schroeder, *ibid.* **58**, 728 (1975)
11. The existence of higher-frequency combination tones can be demonstrated through forced-choice detection of an interaural phase shift between a probe tone and a combination tone. To do this it is necessary to present the probe at the approximately "correct" amplitude and phase. We have used the estimates of amplitude and phase and required subjects to detect a change of 45° in the phase of the contralateral tone at the frequency of a given combination tone. This could be done quite easily.
12. R. M. Sachs and P. M. Zurek, *J. Acoust. Soc. Am.* **62** (Suppl.), 97 (1977).
13. As with other procedures for the measurement of combination-tone amplitude, caution must be exercised in interpreting the obtained probe amplitude directly as the amplitude of the combination tone. As D. D. Greenwood [*J. Acoust. Soc. Am.* **52**, 1155 (1972)] has pointed out, any frequency-dependent transfer function, such as that of the middle ear, intervening before the site of the nonlinear distortion may attenuate the primary tones and probe tone differently. For cases in which the absolute threshold at the frequency of the combination tone is higher than at the primary frequencies, direct interpretation of probe level as combination-tone level would overestimate the cochlear amplitude of the combination tone. When the absolute threshold at the frequency of the combination tone is lower than the primary frequencies, as can be the case with combination tones higher in frequency than the primary tones, the probe level may underestimate the relative cochlear amplitude of the combination tone.
14. S. S. Stevens and H. Davis, *Hearing, Its Psychology and Physiology* (Wiley, New York, 1938), chap. 7; E. G. Weaver and M. Lawrence, *Physiological Acoustics* (Princeton Univ. Press, Princeton, N.J., 1954), ch. 9; P. Dallos, *The Auditory Periphery* (Academic Press, New York, 1973), chap. 6.
15. Supported by grant NS 03856 from the National Institute of Neurological and Communicative Disorders and Stroke to Central Institute for the Deaf. We thank several of our colleagues for criticizing a draft of this report.

13 March 1979

Human Growth Hormone: Complementary DNA Cloning and Expression in Bacteria

Abstract. *The nucleotide sequence of a DNA complementary to human growth hormone messenger RNA was cloned; it contains 29 nucleotides in its 5′ untranslated region, the 651 nucleotides coding for the prehormone, and the entire 3′ untranslated region (108 nucleotides). The data reported predict the previously unknown sequence of the signal peptide of human growth hormone and, by comparison with the previously determined sequences of rat growth hormone and human chorionic somatomammotropin, strengthens the hypothesis that these genes evolved by gene duplication from a common ancestral sequence. The human growth hormone gene sequences have been linked in phase to a fragment of the* trp *D gene of* Escherichia coli *in a plasmid vehicle, and a fusion protein is synthesized at high level (approximately 3 percent of bacterial protein) under the control of the regulatory region of the* trp *operon. This fusion protein (70 percent of whose amino acids are coded for by the human growth hormone gene) reacts specifically with antibodies to human growth hormone and is stable in* E. coli.

Growth hormone, along with at least two other polypeptide hormones, chorionic somatomammotropin (placental lactogen) and prolactin, forms a set of proteins with amino acid sequence homology and to some extent overlapping biological activities (*1*, *2*). Since the genes of this set of proteins probably have a common ancestral origin (*1*), they constitute an excellent model to study the evolution, structure, and differential regulation of related genes. In addition, since human growth hormone is of considerable medical importance and its supply is limited, the synthesis of growth hormone in bacteria might provide the required alternate source of this critical hormone.

We have previously isolated and analyzed bacterial clones containing copies of complementary DNA (cDNA) transcripts of messenger RNA's (mRNA's) for these hormones. The complete sequence of rat pregrowth hormone mRNA (*3*) has been reported; in addition, sequence data have been presented for fragments of about 550 bases complementary to part of the coding (amino acid residues 24 to 191) and 3′ untranslated portions of human chorionic somatomammotropin (hCS) (*3*, *4*) and human growth hormone (hGH) mRNA's (*5*). A partial sequence of rat prolactin has been determined by Gubbins *et al.* (*6*). These sequence data showed that, whereas the growth hormone genes of the rat and man had significant homology, they also had diverged substantially, such that they differed more than the genes for the functionally distinct human hormones hCS and hGH.

We now report the synthesis, cloning, and sequence analysis of cDNA containing the entire coding and most of the noncoding portions of hGH mRNA. We also describe the insertion of these sequences into an "expression plasmid" containing part of the *Escherichia coli* tryptophan (*trp*) operon whose construction has been realized by Hallewell and Emtage (*7*). We describe the use of this plasmid to promote the inducible bacterial synthesis of high levels of a hybrid protein, 70 percent of which is composed of amino acids coded for by the hGH gene.

Human growth hormone mRNA isolation. Polyadenylated RNA was isolated (*8*) from human pituitary tumors removed by transphenoidal hypophysectomy. To obtain an indication of the integrity and the relative abundance of growth hormone mRNA in each sample, the individual mRNA preparations were translated in the wheat germ cell-free system, and the products were analyzed by electrophoresis on sodium dodecyl sulfate–polyacrylamide gels (Fig. 1). Among the translation products of the five acromegalic tumor RNA's (Fig. 1, lanes 1 to 5), the most prominent band corresponds to a protein of approximate-

0036-8075/79/0810-0602$00.50/0

ly 24,000 daltons. This protein is assumed to be human pregrowth hormone since it is similar in size to rat pregrowth hormone (Fig. 1, "rat") and is precipitated by antiserum to hGH (data not shown). This assumption is further justified by comparison with the translation products of polyadenylated RNA isolated from bovine pituitary (Fig. 1, "cow") and from a human prolactin-producing tumor (Fig. 1, lane 6). Both of these RNA's directed the synthesis of a protein similar in size to human and rat pregrowth hormone, but also directed the synthesis of a larger quantity of a protein of higher molecular weight, presumably preprolactin. The tumors show variation in the extent to which hGH mRNA is present, as measured by their translational activities. Nevertheless, hGH mRNA appears to be the most abundant mRNA species in the acromegalic tumors (Fig. 1, lanes 1 to 5). These results are consistent with and verify the clinical diagnoses made prior to surgery.

Molecular cloning of hGH cDNA. The polyadenylated RNA from the tumors that appeared to have the greatest abundance of hGH mRNA by the translational assay were pooled for synthesis of double-stranded cDNA (Fig. 2). Portions of the double-stranded cDNA were analyzed by restriction endonuclease digestion before and after treatment with S1 nuclease. A high proportion of the cDNA was about 1000 nucleotides long (Fig. 2, lanes a and b), the length expected for hGH mRNA, assuming analogy with rat pregrowth hormone mRNA (*3*). Endonuclease Hae III digestion of the DNA generated a 550-base pair (bp) fragment (Fig. 2, lanes h and j) previously reported to occur in hGH and hCS cDNA's (*4, 5*). The prominence of this band supports the idea that the cDNA is highly enriched in hGH gene sequences. This is further suggested by the finding of a fragment of about 400 bp generated by digestion with Hinf I and Sma I (Fig. 2, lane g). The fragment of about 500 bp generated by Pvu II (Fig. 2, lanes e and i) extends beyond the previously cloned 550 bp fragment, which contains only one Pvu II site. However, its presence is predictable from the sequence of rat growth hormone cDNA (*3*), and by the conservation between species of the amino acid sequence in this region (*9*). The fragments of about 350 bp and 150 bp generated by combined digestion with Pvu II and Bgl II (Fig. 2, lane f) would also be anticipated from the previously determined structure of the 550 bp hGH fragment and knowledge of the existence of the additional Pvu II site. Therefore, this cDNA preparation appears to be highly enriched in full-length copies of hGH mRNA.

The uncleaved cDNA was cloned in the plasmid pBR322 and *E. coli* χ1776 in a P3 physical containment facility (*10*) by methods similar to those previously described (*3*). Briefly, the cDNA was first treated with S1 nuclease and subsequently with DNA polymerase I in the presence of the four deoxynucleoside triphosphates to generate blunt-ended cDNA molecules. Synthetic DNA containing the site for the restriction endonuclease Hind III was then added to

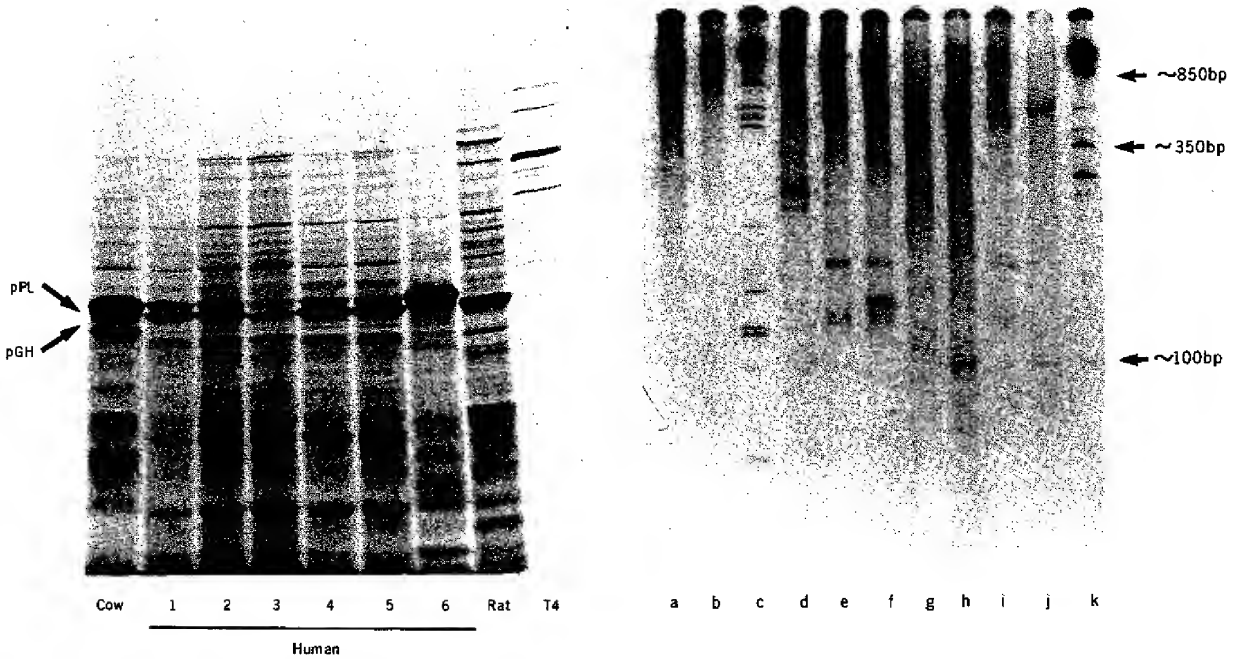


Fig. 1 (left). Translation products of mRNA isolated from growth hormone and prolactin-producing tumors and from bovine pituitary. These tissues were stored in liquid nitrogen shortly after their removal until preparation of the RNA. A portion of polyadenylated RNA isolated from each human tumor, the bovine pituitary, and the cultured rat pituitary tumor (GC) cells was used as a messenger in the wheat germ cell-free protein synthesis system (*20*). The ^{35}S-labeled proteins were analyzed by electrophoresis and autoradiography on sodium dodecyl sulfate-polyacrylamide gels (12.5 percent) (*20*). The translation products from five acromegalic and one prolactin-secreting tumor are shown in lanes 1 to 5 and lane 6, respectively. The lanes labeled "cow" and "rat" show the translation products from bovine pituitary and rat GC cell RNA. Lane 7 shows the bacteriophage T4 proteins (*21*) used as molecular weight markers. The arrows indicate the bands corresponding to pregrowth hormone (pGH) and preprolactin (pPL). Fig. 2 (right). Analysis of cDNA synthesized from mRNA extracted from the growth hormone-producing pituitary tumors. The polyadenylated RNA from tumors two, four, and five were pooled (135 μg) and used as a template for synthesis of ^{32}P-labeled double-stranded cDNA as described (*3*). Samples of this cDNA were cleaved with various restriction endonucleases before and after S1 nuclease digestion. The figure shows an autoradiogram of the resulting DNA fragments after electrophoresis on a 4.5 percent polyacrylamide gel (*3*). (Lane a) Uncleaved cDNA before S1 digestion; (lane b) uncleaved cDNA after S1 digestion; (lane c) bacteriophage fd DNA, Hpa II (molecular weight markers); (lane d) cDNA, Pst I + Bgl II; (lane e) cDNA, Pvu II; (lane f) cDNA, Bgl II + Pvu II; (lane g) cDNA, Hinf I + Sma I; (lane h) cDNA, Hae III; (lane i) cDNA, Pvu II; (lane j) cDNA, Hae III; (lane k) bacteriophage fd DNA, Hae III (molecular weight markers).

each end of the cDNA, and cohesive ends were generated by digestion with endonuclease Hind III. The resulting cDNA was purified on gel and ligated to Hind III–cut and bacterial alkaline phosphatase–treated pBR322 plasmid DNA (*8*). Bacteria were then transformed with this recombinant plasmid. Colonies with recombinant DNA containing plasmids, selected by antibiotic resistance ($tet^S amp^R$) were grown, and the plasmid DNA's were isolated. The DNA was digested with Hind III, treated with various other restriction endonucleases, and analyzed by gel electrophoresis. One clone contained an insert of about 800 bp whose digestion by Hae III, Pvu II, and Bgl II generated fragments similar in size

```
                                        -26                 -20                 -10
                                        met ala thr gly ser arg thr ser leu leu leu ala phe gly leu leu cys leu pro trp
GG AUC CUG UGG ACA GCU CAC CUA CCU CCA AUG GCU ACA CGC UCC CGC ACG UCC CUG CUC CUG GCU UUU GGC CUC CUC UGC CUC CCC UGC

                        1                   10                  20
leu gln glu gly ser ala phe pro thr ile pro leu ser arg leu phe asp asn ala met leu arg ala his arg leu his gln leu ala
CUU CAA CAC GCC AGU GCC UUC CCA ACC AUU CCC UUA UCC AGC CUU UUU CAC AAC GCU AUG CUC CGC GCC CAU CGU CUG CAC CAG CUG GCC

                        30                  40                  50
phe asp thr tyr gln glu phe glu glu ala tyr ile pro lys glu gln lys tyr ser phe leu gln asn pro gln thr ser leu cys phe
UUU GAC ACC UAC CAG GAG UUU GAA GAA GCC UAU AUC CCA AAG GAA CAG AAG UAU UCA UUC CUG CAG AAC CCC CAG ACC UCC CUC UGU UUC

                        60                  70                  80
ser glu ser ile pro thr pro ser asn arg glu glu thr gln gln lys ser asn leu glu leu leu arg ile ser leu leu leu ile gln
UCA GAG UCU AUU CCG ACA CCC UCC AAC AGG GAG GAA ACA CAA CAG AAA UCC AAC CUA GAG CUG CUC CGC AUC UCC CUG CUC CUC AUC CAG

                        90                  100                 110
ser trp leu glu pro val gln phe leu arg ser val phe ala asn ser leu val tyr gly ala ser asp ser asn val tyr asp leu leu
UCG UGG CUG GAG CCC GUG CAG UUC CUC AGC AGU GUC UUC GCC AAC AGC CUG GUC UAC GGC GCC UCU GAC AGC AAC GUC UAU GAC CUC CUA

                        120                 130                 140
lys asp leu glu glu gly ile gln thr leu met gly arg leu glu asp gly ser pro arg thr gly gln ile phe lys gln thr tyr ser
AAG GAC CUA GAG GAA GGC AUC CAA ACG CUG AUG GGG AGG CUC GAA GAU GGC ACC CCC CGG ACU GGG CAG AUC UUC AAA CAG ACC UAC AGC

                        150                 160                 170
lys phe asp thr asn ser his asn asp asp ala leu leu lys asn tyr gly leu leu tyr cys phe arg lys asp met asp lys val glu
AAG UUC GAC ACA AAC UCA CAC AAC GAU GAC GCA CUA CUC AAG AAC UAC GGG CUG CUC UAC UGC UUC AGG AAG GAC AUG GAC AAG GUC GAG

                        180                 190 191
thr phe leu arg ile val gln cys arg ser val glu gly ser cys gly phe AM
ACA UUC CUG CGC AUC GUG CAG UGC CGC UCU GUG GAG GCC AGC UGU GGC UUC UAG CUG CCC GGC UGG CAU CCU GUG ACC CCU CCC CAG UGC

CUC UCC UGG CCC UGG AAG UUC CCA CUC CAC UCC CCA CCA GCC UUC UCC UAA UAA AAU UAA GUU GCA UCA AAA AAA AAA
```

Fig. 3. Nucleotide sequence of hGH mRNA and the amino acid sequence of human pregrowth hormone. The sequence was determined according to the procedure of Maxam and Gilbert (*11*) from 5'- or 3'-end-labeled restriction fragments of chGH800/pBR322. The sequence between the two internal Hae III sites was taken from our previous work (*5*). Most of it, and all of the sequence outside these two internal Hae III sites was resequenced by the chain–termination technique (*12*) as described in detail elsewhere (*22*) using as single-stranded templates the chGH800 Hind III fragment recloned in the vector M13 mp5 and as primers restriction fragments of chGH800/pBR322. The RNA sequence has been taken from the DNA sequence. The amino acid sequence has been deduced from the RNA sequence using the genetic code. The termination codon, UAG, is designated by the symbol AM for "amber."

4/UUU/phe	3/UCU/ser	3/UAU/tyr	2/UGU/cys
10/UUC/phe	7/UCC/ser	5/UAC/tyr	3/UCC/cys
1/UUA/leu	3/UCA/ser	0/UAA/OC	0/UGA/OP
0/UUG/leu	1/UCG/ser	1/UAG/AM	2/UGG/trp
2/CUU/leu	0/CCU/pro	1/CAU/his	1/CGU/arg
10/CUC/leu	6/CCC/pro	2/CAC/his	4/CGC/arg
4/CUA/leu	2/CCA/pro	3/CAA/gln	0/CGA/arg
16/CUG/leu	1/CCG/pro	11/CAG/gln	2/CCG/arg
2/AUU/ile	1/ACU/thr	0/AAU/asn	2/AGU/ser
6/AUC/ile	4/ACC/thr	9/AAC/asn	5/AGC/ser
0/AUA/ile	5/ACA/thr	1/AAA/lys	0/AGA/arg
4/AUG/met	2/ACG/thr	8/AAG/lys	5/AGG/arg
0/GUU/val	3/GCU/ala	2/GAU/asp	0/GGU/gly
3/GUC/val	6/GCC/ala	9/GAC/asp	8/GGC/gly
0/GUA/val	1/GCA/ala	6/GAA/glu	0/GGA/gly
4/GUG/val	0/GCG/ala	9/GAG/glu	3/GGG/gly

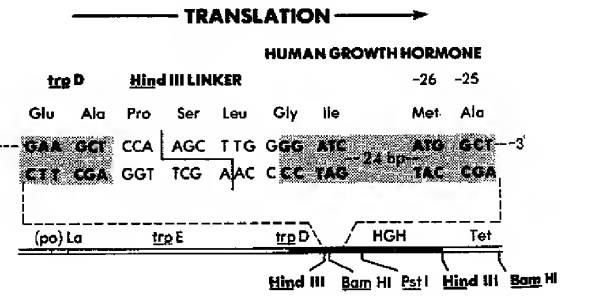


Fig. 4. Codon usage in hGH mRNA. The numbers indicate how many times the codons are used in the region of hGH mRNA coding for the prehormone; OC, OP, and AM designate the stop codons ochre, opal, and amber, respectively. Fig. 5. Postulated nucleotide sequence around the Hind III site in the hybrid gene of expression plasmid p*trp*ED50-chGH800. Plasmid p*trp*ED50 was constructed from p*trp*ED5-1 (*7*) by linearizing the plasmid with Hind III, filling in the protruding 5' ends with the use of DNA polymerase I (Klenow fragment from Boehringer) and ligating synthetic decamers containing a Hind III site (collaborative research) to the blunt-ended material (*3*, *8*). After digestion with Hind III the plasmid was separated from residual linker molecules by chromatography on Sephadex G-200, recircularized with T4 DNA ligase, and used to transform *E. coli* W3110 trpoE▼1(*23*) by a standard procedure (*24*). Plasmid DNA isolated from one of these colonies was digested with Hind III and treated with alkaline phosphatase (*4*, *8*). A portion (5 μg) of this DNA was end-labeled with [γ-^{32}P]-ATP with the use of T4 DNA kinase (Boehringer) and cut by Hae III. The DNA sequence of the labeled fragments was determined by chemical cleavage (*11*), after they were isolated by polyacrylamide electrophoresis. The cloned chGH800 DNA was cleaved from pBR322 with Hind III and isolated by polyacrylamide gel electrophoresis. This DNA was ligated to similarly cleaved and alkaline phosphatase-treated p*trp*ED50. The ligation mixture was used to transform *E. coli* strains W3110 *trp*oE▼1 and RR1 (*25*) in a P3 facility, and transformants resistant to ampicillin were selected. Resistant colonies were examined for the presence of inserted chGH800 sequences by gel analysis of such plasmids after digestion with restriction endonucleases Bam HI and Pst I. Plasmids with the growth hormone initiator codon proximal to the *trp*D gene sequence showed bands of 250 and 900 bp, whereas plasmids with the inserted cDNA in the opposite orientation showed bands of 250 and 350 bp.

to those from the 550 bp hGH cDNA clone (data not shown) and to the digested uncloned cDNA (Fig. 2). This suggested that this clone did contain cDNA complementary to full-length or nearly full-length hGH mRNA. (This clone is designated chGH800/pBR322.)

Sequence analysis of cloned DNA. The nucleotide sequence of the cloned DNA was determined by the chemical cleavage method of Maxam and Gilbert (*11*) and the chain-termination technique of Sanger, Nicklen, and Coulson (*12*). The hGH mRNA sequence and the corresponding amino acid sequence of human pregrowth hormone can be derived from the DNA sequence (Fig. 3). The amino acid sequence determined from the DNA sequence is consistent with the known amino acid sequence of hGH (*13*) with the following exceptions: the DNA sequence predicts glutamine, asparagine, glutamine, glutamic acid, glutamine, aspartic acid, asparagine, and glutamine at amino acid positions 29, 47, 49, 74, 91, 107, 109, and 122, respectively, while the protein sequence indicates glutamic acid, aspartic acid, glutamic acid, glutamine, glutamic acid, asparagine, aspartic acid, and glutamic acid. It is likely that the DNA sequence is correct in this regard since it is sometimes difficult in protein sequence analysis to differentiate aspartic acid from asparagine and glutamic acid from glutamine. The amino acid sequence of the signal peptide portion of human pregrowth hormone had not been previously determined and is deduced from the mRNA sequence. If translation begins with the methionine codon "in phase," 26 codons proximal to the first amino acid of growth hormone (Fig. 3), then the primary translation product of hGH mRNA would be a protein of 24,851 daltons, a value in agreement with the cell-free translation data shown in Fig. 1.

A comparison of the amino acid and nucleic acid sequence homologies between rat growth hormone, hGH, and hCS and their respective mRNA's is shown in Table 1. In the coding regions, there is higher homology between the nucleic acid sequences than between the amino acid sequences. This difference is consistent with the already mentioned view that the genes of these related hormones evolved from a common evolutionary precursor gene, and is further supported by the marked homology in the 5′-noncoding portions of the mRNA's for rat and human growth hormone. (Data for the 5′-noncoding region of hCS are not yet available.) Human growth hormone has more homology with hCS than with rat growth hormone, especially for the 3′-noncoding portions. This finding supports the hypothesis developed earlier that the chorionic somatomammotropin and growth hormone genes probably evolved by a gene duplication mechanism (*1*) at some time after the separation of the human and rat species. In addition, the fact that both hormones exist in both species implies that the same hormones may have evolved independently more than once.

Table 1. Amino acid and nucleic acid sequence homology of growth hormone, chorionic somatomammotropin, and their mRNA's. Data for rat growth hormone (rGH) and human chorionic somatomammotropin (hCS) and their mRNA's are from (*3*, *4*). For hCS, only data for amino acids residues 24 to 191 (and the corresponding portion of the mRNA) and a portion of the noncoding 3′-region corresponding to the cloned 550 bp fragment are used for comparisons, since data for the other portions are not available. The amino acid sequence of the prepeptide portion of hCS was determined by Sherwood *et al.* (*27*).

Source	Homology (percent)	
	hGH versus rGH	hGH versus hCS
Nucleic acid:		
5′-Noncoding	73	
Presequence	76	
Coding	76	92
3′-Noncoding	38*	94
Amino acid:		
Presequence	58	84
Coding	67	86

*The homology in this region can be increased to 55 percent by adding appropriate gaps in the sequences. This procedure also reveals that there is a homology of 27 out of 30 bases in the region of the AAUAAA (*28*). Similar conservation between species but with different sequences are found when human and rabbit β-globin (*29*) and human and rat insulin (*30*) are compared.

Figure 4 shows the codons used for hGH mRNA. As is the case with rat growth hormone (*3*) and hCS (*6*) mRNA's, there is a nonrandom selection of codons. This appears to be mostly due to the preference for G (guanine) or C (cytosine) over A (adenine) or U (uracil) for the third position of the triplet codon. This is also the case with most (*3*, *14*) but not all (*14*) eukaryotic mRNA's whose structures are known.

Construction of a plasmid for growth hormone expression. To see whether hGH gene sequences can be expressed in bacteria, we used the plasmid ptrpED5-1 (*7*), which contains the regulatory region [(po)La], the first gene (*trp*E), and 15 percent of the second gene (*trp*D) of the *E. coli trp* operon. Cells containing ptrpED5-1 normally synthesize small amounts of *trp* gene products. However, if *trp* operon transcription is derepressed by addition of 3β-indolylacrylic acid, synthesis of *trp* gene products increases, so that within 3 hours *trp* proteins account for about 30 percent of the total cellular proteins (*7*). We hoped that by placing the hGH gene sequence under control of the *trp* operon, not only would it be expressed, but a higher level of hGH production could be obtained than was previously achieved with rat growth hormone gene sequences under control of the β-lactamase gene (*15*).

The hGH sequences from chGH800 were inserted at the Hind III site of the *trp*D gene sequence as described in the legend to Fig. 5. In order to insert the hGH codons in phase with those of the *trp*D sequence, the Hind III site in the *trp*D gene was manipulated in such a way as to shift the reading frame of any DNA inserted through the Hind III site of the plasmid by one base. To do this, ptrpED5-1 was cleaved with Hind III, the protruding 5′-ends "filled in" with the use of DNA polymerase I (Klenow fragment) and synthetic DNA decamers containing the Hind III site ligated to the blunt-ended material. This DNA was then digested with Hind III to produce new cohesive Hind III ends, and plasmid molecules were recircularized with the use of DNA ligase after the residual Hind III linker molecules were removed on a Sephadex (G-200) column. This material was used to transform *E. coli* and, after selection for ampicillin-resistant colonies, the plasmid DNA was isolated from one of the transformants. The DNA sequence at the Hind III site of the newly constructed plasmid (designated ptrpED50) was determined and showed that the enzymatic reactions had altered the reading frame as predicted. In this way, when chGH800/pBR322 was cleaved with Hind III and the hGH sequences ligated to the Hind III site of the newly constructed plasmid, the codons of hGH would be in phase with those of the *trp*D gene, provided that the hGH gene sequences were inserted in the proper orientation. This was achieved by obtaining several clones, isolating plasmid DNA, and determining the orientation of the cloned segment by restriction endonuclease analysis (Fig. 5, legend).

As is indicated in Fig. 5, the construction of the hGH "expression" plasmid was such that the anticipated product would be a fusion protein containing the NH_2-terminal region of the *trp*D protein, amino acids coded by the 5′-untranslated portion of hGH mRNA, the 26 amino acids of the signal peptide, and all of the amino acids of hGH.

Synthesis of growth hormone in bacteria. To determine whether the newly

constructed gene can direct the synthesis of large amounts of a new fused polypeptide, and whether its expression is regulated by the *trp* promoter, cells containing the expression plasmid were derepressed for *trp* transcription, and proteins were labeled for 5 minutes with ^{14}C-labeled amino acids. Figure 6 shows an autoradiogram of sodium dodecyl sulfate–polyacrylamide gel of such proteins labeled at various times from 0 to 4 hours after induction of the *trp* operon. Two proteins (53,000 and 32,000 daltons) seem to be specifically derepressed by the inducer. The higher molecular weight protein is the *trp*E gene product (*7*). The 32,000-dalton protein has approximately the anticipated size for the *trp*D–hGH fusion protein (34,000 daltons). It is immunoprecipitated by antiserum to hGH (Fig. 6, lane a) but not by the control antiserum (Fig. 6, lane b), and precipitation can be blocked by a large excess of hGH (Fig. 6, lanes c to h). Some of the *trp*E protein is immunoprecipitated by antiserum to hGH, but the amount is less when an excess of competitor hGH is added or control antiserum is used. This result might be expected for two reasons. Precipitation of *trp*E protein by control antiserum may be due to the high abundance of this protein. More interestingly, specific precipitation of *trp*E by hGH antiserum (Fig. 6, lane c) and blockage of precipitation by an excess of competitor hGH (Fig. 6, lane e) may be the result of association of the *trp*E protein and the *trp*D–hGH fused polypeptide. The *trp*E and *trp*D proteins are normally associated in *E. coli* as a tetramer containing two subunits of each protein (*16*); the resulting enzymatic activity (anthranilate synthetase) requires the *trp*E protein and the NH_2-terminal 30 percent of the *trp*D protein (*17*). Thus, the fused *trp*D–hGH protein may contain those *trp*D residues required for binding *trp*E. All of these lines of evidence suggest that the 32,000-dalton protein is a fused *trp*D–hGH polypeptide.

On the basis of the relative quantity of radioactivity incorporated into the *trp*D–hGH gene product, the fusion product appears to be a major protein made by the bacteria, constituting 3 percent of the total bacterial protein synthesis. Thus, the natural hGH gene sequence can be expressed at a high level in bacteria.

The *trp*E and *trp*D protein molecules ordinarily accumulate at a similar rate (molar ratio 1:1) when the *trp* operon is induced (*7*). However, the molar ratio of *trp*E to *trp*D-hGH is about 6:1, indicating that synthesis of *trp*D–hGH is only 17 percent of the expected level. The reduced level of synthesis is not due to instability of the fused polypeptide since a "pulse-chase" experiment has shown that no significant degradation of the fused polypeptide occurs during the 60-minute period after incorporation of the label (data not shown). There is some evidence that the chick ovalbumin protein is also synthesized in *E. coli* at lower levels than expected (*18*).

Hypopituitary dwarfism is a fairly common disease treatable only by replacement with hGH (*19*). Growth hormone may also be useful in the treatment of other disorders. However, the potential uses of this hormone have not been adequately investigated because its only source is pituitaries from human cadavers. In order to have an adequate supply of this hormone, it is necessary to find alternative means of producing it; the synthesis of hGH in bacteria may provide such a means.

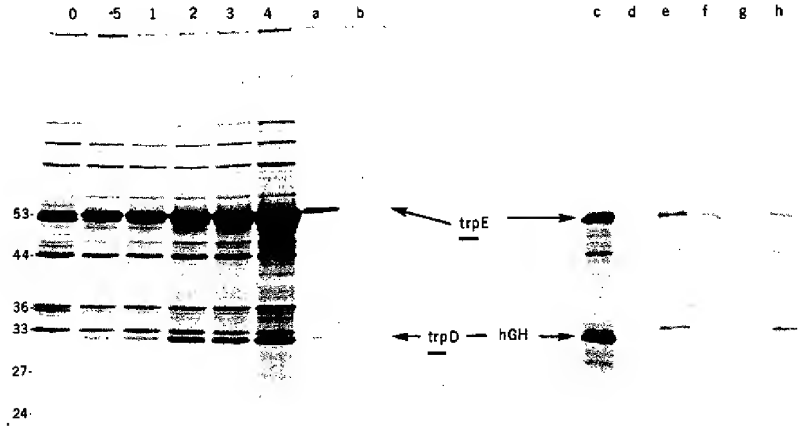


Fig. 6. Autoradiograms of sodium dodecyl sulfate–polyacrylamide gels (10 percent) of ^{14}C- and ^{35}S-labeled proteins from bacteria harboring the p*trp*ED50-chGH800 expression plasmid. Cultures of strains W3110 *trp*oEV1 and RRI harboring this plasmid were induced with 3β-indolylacrylic acid and 3-ml samples labeled for 5 minutes with 2 μCi of ^{14}C-labeled amino acids (W3110 *trp*oEV1) or 10 μCi of ^{35}S-labeled methionine (RRI) as described (*7*). Samples labeled at zero, 0.5, 1, 2, 3, and 4 hours were centrifuged and resuspended (50 μl) by sonication (*26*) prior to loading (5 μl per gel slot) in sodium dodecyl sulfate sample buffer (*20*). Samples were immunoprecipitated by means of the SAC technique (*20*) in order to collect antigen-antibody complexes. Immunoprecipitations contained 10 μl of sonicated cells, 390 μl of 0.5 percent NP40 (Particle Data Laboratory, Elmhurst, Ill.) in phosphate saline buffer (0.025*M* potassium phosphate, *p*H 7.4, 0.1*M* NaCl), 20 μl of rabbit antiserum to hGH (Antibodies Inc., 1000 unit/ml) or 20 μl of nonimmune rabbit antiserum each diluted 50-fold in phosphate saline buffer containing bovine serum albumin (2 mg/ml). Competitor hGH was added at 40 μg per reaction mixture. The *E. coli* proteins in lane 0 were used as molecular weight markers (*8*). (a and b) Immunoprecipitates of ^{14}C-labeled proteins from the 4-hour time point with (a) antiserum against hGH and with (b) nonimmune serum. (c to e) Immunoprecipitates of ^{35}S-labeled proteins from the 4-hour time point, with (c) antiserum against hGH, (d) nonimmune serum, and (e) antiserum against hGH together with an excess of competitor hGH. (f to h) Immunoprecipitates of ^{35}S-labeled proteins from the zero time point with the use of (f) antiserum against hGH, (g) nonimmune serum, and (h) antiserum against hGH together with an excess of competitor hGH.

JOSEPH A. MARTIAL
ROBERT A. HALLEWELL
JOHN D. BAXTER
HOWARD M. GOODMAN
Howard Hughes Medical Institute Laboratories, Department of Medicine, Metabolic Research Unit and *Department of Biochemistry and Biophysics, University of California, San Francisco 94143*

References and Notes

1. T. A. Bewley, J. S. Dixon, C. H. Li, *Int. J. Pept. Protein Res.* **4**, 281 (1972); H. D. Niall, M. L. Hogan, R. Sayer, I. Y. Rosenblum, F. C. Greenwood, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 866 (1971).
2. A. G. Frantz, in *Peptide Hormone*, J. A. Parson, Ed. (University Park Press, Baltimore, 1976), p. 199.
3. P. H. Seeburg, J. Shine, J. A. Martial, J. D. Baxter, H. M. Goodman, *Nature (London)* **270**, 5637 (1977).
4. P. H. Seeburg, J. Shine, J. A. Martial, A. Ullrich, J. D. Baxter, H. M. Goodman, *Cell* **12**, 157 (1977).
5. P. H. Seeburg, J. Shine, J. A. Martial, J. D. Baxter, H. M. Goodman, in preparation; H. M. Goodman, P. H. Seeburg, J. Shine, J. A. Martial, J. D. Baxter, *Alfred Benson Symposium 1978* (Munsgaard, Copenhagen, in press).
6. E. J. Gubbins, R. A. Maurer, J. L. Hartley, J. E. Donelson, *Nucleic Acids Res.* **6**, 915 (1979); N. E. Cooke and J. A. Martial, unpublished results.
7. R. A. Hallewell and S. Emtage, in preparation.
8. A. Ullrich, J. Shine, J. Chirgwin, R. Pictet, E. Tischer, W. J. Rutter, H. M. Goodman, *Science* **196**, 1313 (1977).
9. M. O. Dayhoff, *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, D.C., 1972), vol. 5, pp. D-201–D-204.
10. "Recombinant DNA research, revised guidelines," *Fed. Reg.* **43**, 60080 (1978); "Recombi-

nant DNA research: actions under guidelines," *ibid.* **44**, 21730 (1979).
11. A. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
12. F. Sanger, S. Nicklen, A. R. Coulson, *ibid.*, p. 5463.
13. Reference (*9*), p. D-202, was used as the "standard" hGH sequence, except that an additional Gln residue was placed at position 69 as indicated in (*1*) and (*2*). References (*1*) and (*2*) present slightly different versions of the hGH sequence, which match the sequence predicted from the cDNA in several residues, which do not match the standard sequence.
14. L. McReynolds, B. W. O'Malley, A. D. Nisbet, J. E. Fothergill, D. Givol, S. Fields, M. Robertson, G. G. Brownlee, *Nature (London)* **273**, 723 (1978).
15. P. H. Seeburg, J. Shine, J. A. Martial, R. D. Ivarie, J. A. Morris, A. Ullrich, J. D. Baxter, H. M. Goodman, *ibid.* **276**, 5690 (1978).
16. J. Ito and C. Yanofsky, *J. Bacteriol.* **97**, 734 (1969).
17. C. Yanofsky, V. Horn, M. Bonner, S. Stasiowski, *Genetics* **69**, 409 (1971).
18. T. H. Fraser and B. J. Bruce, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5936 (1978).
19. J. M. Tanner, *Nature (London)* **237**, 433 (1972).
20. J. A. Martial, J. D. Baxter, H. M. Goodman, P. H. Seeburg, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 1816 (1977).
21. P. A. O'Farrell and L. H. Gold, *J. Biol. Chem.* **248**, 7066 (1973).
22. B. Cordell, G. Bell, E. Tischer, F. M. DeNoto, A. Ullrich, R. Pictet, W. J. Rutter, H. M. Goodman, in preparation.
23. N. E. Murray and W. J. Brammar, *J. Mol. Biol.* **77**, 615 (1973).
24. D. M. Glover, in *New Techniques in Biophysics and Cell Biology*, R. H. Pain and B. J. Smith, Eds. (Wiley, New York, 1976), vol. 8, pp. 125–145.
25. F. Bolivar, R. L. Rodriguez, M. C. Betlach, H. W. Boyer, *Gene* **2**, 75 (1977).
26. P. H. O'Farrell, *J. Biol. Chem.* **250**, 4007 (1975).
27. L. M. Sherwood, Y. Burstein, I. Schechter, *Conference on Precursor Processing in The Biosynthesis of Proteins* (New York Academy of Sciences, 2 to 4 May 1979), Abstr., S. Birken, D. L. Smith, R. E. Canfield, I. Boime, *Biochem. Biophys. Res. Commun.* **74**, 106 (1977).
28. N. J. Proudfoot and G. G. Brownlee, *Nature (London)* **263**, 211 (1976).
29. A. Efstratiadis, F. C. Kafatos, T. Maniatis, *Cell* **10**, 571 (1977).
30. G. I. Bell, W. F. Swain, R. Pictet, B. Cordell, H. M. Goodman, W. J. Rutter, *Nature (London)*, in press.
31. Supported by a grant from Eli Lilly Co. and a postdoctoral fellowship from the British Science Research Council (to R.A.H.). We thank Dr. C. Yanofsky for communicating information prior to publication and for helpful suggestions on the manuscript; J. Messing for M13mp5 and advice on its use; Drs. W. Swain and P. O'Farrell for their respective gifts of plasmid pBR322 and radioactive T4 protein; Drs. P. Seeburg and J. Shine for helpful discussion and participation in some experiments; and D. Coit and E. Tischer for technical assistance. Avian myeloblastosis virus reverse transcriptase was provided by the Office of Program Resources and Logistics, NCI. J.D.B. and H.M.G. are investigators of the Howard Hughes Medical Institute.

8 June 1979

Changed Lyotropic Liquid Crystalline Structure Due to Polymerization of the Amphiphilic Component

Abstract. *Optical patterns in polarized light and x-ray reflections in the low-angle region were used to detect a shift from one liquid crystalline structure to another during polymerization. The polymerization took place in a lyotropic liquid crystal of water and sodium undecenoate, with a structure consisting of cylinders in a two-dimensional hexagonal close packing. After polymerization, a lamellar liquid crystalline structure was obtained.*

The comprehensive knowledge of liquid crystalline structures (*1–4*) that has been obtained through their use for display systems and the excellent properties of Kevlar fibers drawn from liquid crystals (*5*) have resulted in a recent focus on polymerization in systems of liquid crystalline character. The contributions so far have mainly been in the area of thermotropic liquid crystals (*6*); the equilibrium problems being more complex in lyotropic structures, which by definition are multicomponent systems.

Early attempts to polymerize in lyotropic liquid crystals (*7–10*) have recently (*11*, *12*) been criticized as achieving retention of the long-range order by freezing the structure with cross-coupling agents rather than forming an equilibrium liquid crystal. The conditions for polymerization with retained liquid crystalline structure have been stated (*11*) by comparison with homotropic polymerization in crystalline structures (*12–15*). For a liquid crystalline structure the translational entropy component in the expression for the total free energy should also be included. Its magnitude is difficult to estimate; for microemulsions (*16*) the contribution of the entropic free energy is similar in magnitude to the contributions of other components of the free energy.

The fact that polymerization with retention of the liquid crystalline structure may be a rare phenomenon in lyotropic liquid crystals (*11*) encouraged us to choose an alternative route: to observe structural changes during polymerization in a lyotropic liquid crystalline matrix. We are now able to report a change from one liquid crystalline structure to another during polymerization of the amphiphilic component. To our knowledge, this is the first report of such a change. It is essential to realize that the structure obtained represents the stable conformation of the system; there is no "freezing in" of a structure—it forms spontaneously from an isotropic melt.

The components of the liquid crystal were distilled water (with 0.05*M* ammonium persulfate as an initiator) and sodium undecenoate. They formed a liquid crystal in the concentration range 47 to 59 percent sodium undecenoate (by

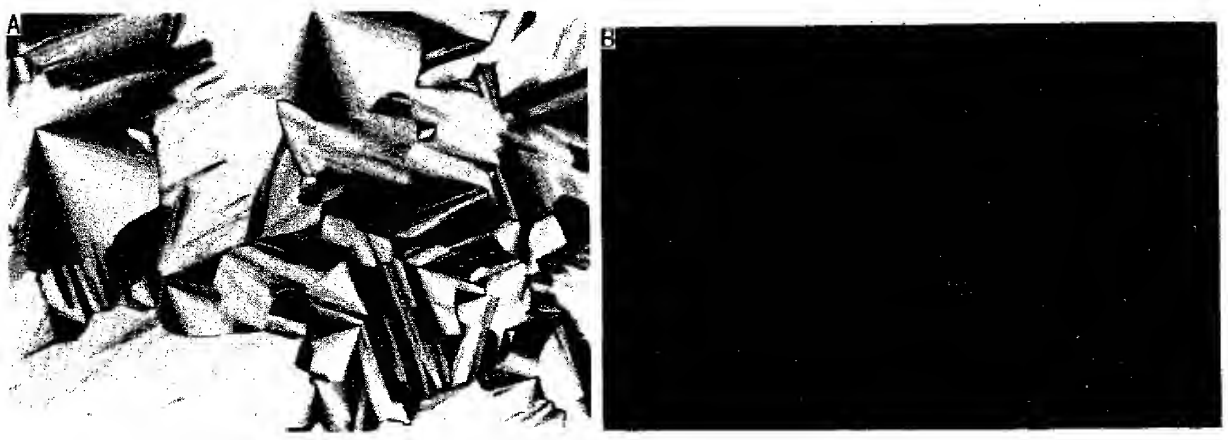


Fig. 1. (A) Optical pattern of the liquid crystalline phase before polymerization, typical of a structure of hexagonally close-packed cylinders. (B) Optical pattern of the polymerized sample.

0036-8075/79/0810-0607$00.50/0

TAB U

Nature Vol. 281 18 October 1979

Direct expression in *Escherichia coli* of a DNA sequence coding for human growth hormone

David V. Goeddel*, Herbert L. Heyneker*, Toyohara Hozumi†, Rene Arentzen†, Keiichi Itakura†, Daniel G. Yansura*, Michael J. Ross*, Giuseppe Miozzari*, Roberto Crea‡ & Peter H. Seeburg*

Divisions of Molecular Biology* and Organic Chemistry‡, Genentech Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080
† Division of Biology, City of Hope National Medical Center, Duarte, California 91010

DNA coding for human growth hormone was constructed by using chemically synthesised DNA in conjunction with enzymatically prepared cDNA. This 'hybrid' gene was expressed in Escherichia coli *under the control of the* lac *promoter. A polypeptide was produced having the size and immunological properties characteristic of mature human growth hormone.*

HUMAN GROWTH HORMONE (HGH) is a protein of 191 amino acids which is synthesised in the anterior lobe of the pituitary. Growth in hypopituitary dwarfs, whose small stature is due to a deficiency of HGH, can be restored during childhood by administration of this hormone[1]. In addition, HGH may prove effective in the treatment of a variety of ailments, including bone fractures, skin burns and bleeding ulcers[2]. As growth hormone is species specific, human cadavers have been the only source of HGH.

The primary translation product of growth hormone mRNA is a precursor protein consisting of a signal peptide attached to the N-terminus of growth hormone[3]. Such signal or 'pre' sequences are characteristic of secreted proteins. In the case of rat growth hormone (RGH), DNA sequencing of cDNA prepared from RGH mRNA has identified the 26 amino acid residues of the pre-sequence[4]. From information derived from the cDNA sequence of human growth hormone[5] we have designed and constructed a bacterial plasmid which instructs the synthesis of substantial quantities of mature HGH in a microbial cell.

HGH gene assembly strategy

The general approach used for the bacterial synthesis of HGH involves a combination of the cloning of the complementary DNA (cDNA) prepared from pituitary mRNA with the cloning of chemically synthesised DNA. The specific strategy was based on the known restriction endonuclease pattern of HGH cDNA[5]. *Hae*III sites are present in the 3' noncoding region and in the sequence coding for amino acids 23 and 24 of HGH. Treatment of double stranded (ds) HGH cDNA with *Hae*III gives a DNA fragment of 551 base pairs which includes coding sequences for amino acids 24–191 of HGH. We planned to clone this cDNA fragment and to clone separately a chemically synthesised DNA 'adaptor' fragment containing an ATG initiation codon and coding sequences for residues 1–23 of HGH. These two DNA fragments would be combined to form a synthetic–natural 'hybrid' gene. When inserted into a plasmid downstream from a suitable bacterial promoter and ribosome binding site, this gene could be expected to direct the synthesis of fMet-HGH. The fact that most bacterial proteins do not contain N-terminal methionine residues suggests that the fMet should be efficiently removed, resulting in the direct expression of HGH.

There are several advantages of this approach over conventional expression methods which utilise either chemically synthesised DNA or cDNA exclusively. The cDNA approach has been used to express several fused proteins (β-lactamase–rat proinsulin[6], β-lactamase–rat pre-growth hormone[7] and β-galactosidase–chicken ovalbumin[8,9]) in *Escherichia coli* which still retain antigenic activity, but which cannot be easily processed to give the natural gene product. However, the cloning of mouse dihydrofolate reductase cDNA into G-tailed, *Pst*I-cleaved pBR322 resulted in the expression of a protein which appeared to be dihydrofolate reductase[10]. This protein could have been derived from the proteolytic cleavage of a β-lactamase–dihydrofolate reductase fused protein or from the direct initiation of translation at the dihydrofolate reductase ATG (AUG) start codon. Neither of these possibilities could be relied on for HGH expression. *In vivo* proteolytic cleavage of a fused bacterially produced protein to give HGH is highly unlikely, and initiation of translation at an HGH start codon would be expected to yield the pre-hormone containing an extra 26 amino acids. The alternative approach of complete chemical synthesis of genes, although shown to be extremely effective for the bacterial production of the human peptides somatostatin[11] and insulin[12], would be very time consuming for a polypeptide as large as HGH.

Construction and cloning of plasmids containing HGH DNA sequences

The plan for the chemical synthesis of DNA designed to code for the first 24 amino acids of HGH is shown in Fig. 1*a*. To facilitate joining to HGH cDNA, the synthetic plan incorporates the same *Hae*III restriction site at amino acid residues 23 and 24 that is found in the cDNA. The 12 indicated deoxyoligonucleotides (Fig. 1*a*) were synthesised by the improved phosphotriester method essentially as described previously for insulin[13]. Figure 1*b* summarises the steps involved in the assembly and cloning of these DNA fragments in pBR322 (ref. 15). Many transformants of *E. coli* 294 (ref. 16) were obtained with plasmids containing *Eco*RI–*Hin*dIII inserts of approximately the desired size. The nucleotide sequences of three of these inserts were determined. One, pHGH3, had the expected DNA sequence. The other two had a deletion of the A · T base pair at the third nucleotide position for the codon of amino acid 5 (Pro).

The construction of the plasmid (pHGH31) containing the *Hae*III cDNA fragment of HGH is outlined in Fig. 2. Total poly(A)-mRNA from human pituitaries was used in the synthesis of ds-cDNA by RNA-dependent DNA polymerase (reverse transcriptase) and DNA polymerase I Klenow fragment. The cDNA was treated with *Hae*III restriction endonuclease and DNA of about 550 base pairs was purified by gel electrophoresis. This DNA fraction was then 'tailed' with deoxyC residues using terminal deoxynucleotidyl transferase[10] and cloned in pBR322 DNA which had been cleaved with *Pst*I and extended with deoxyG residues. This procedure restores the *Hae*III restriction sites on the cDNA and regenerates *Pst*I sites at both ends of the insert. Approximately 200 transformants of *E. coli* χ1776 (ref. 20) were obtained from 1 ng of C-tailed cDNA. These colonies were screened for the presence of HGH

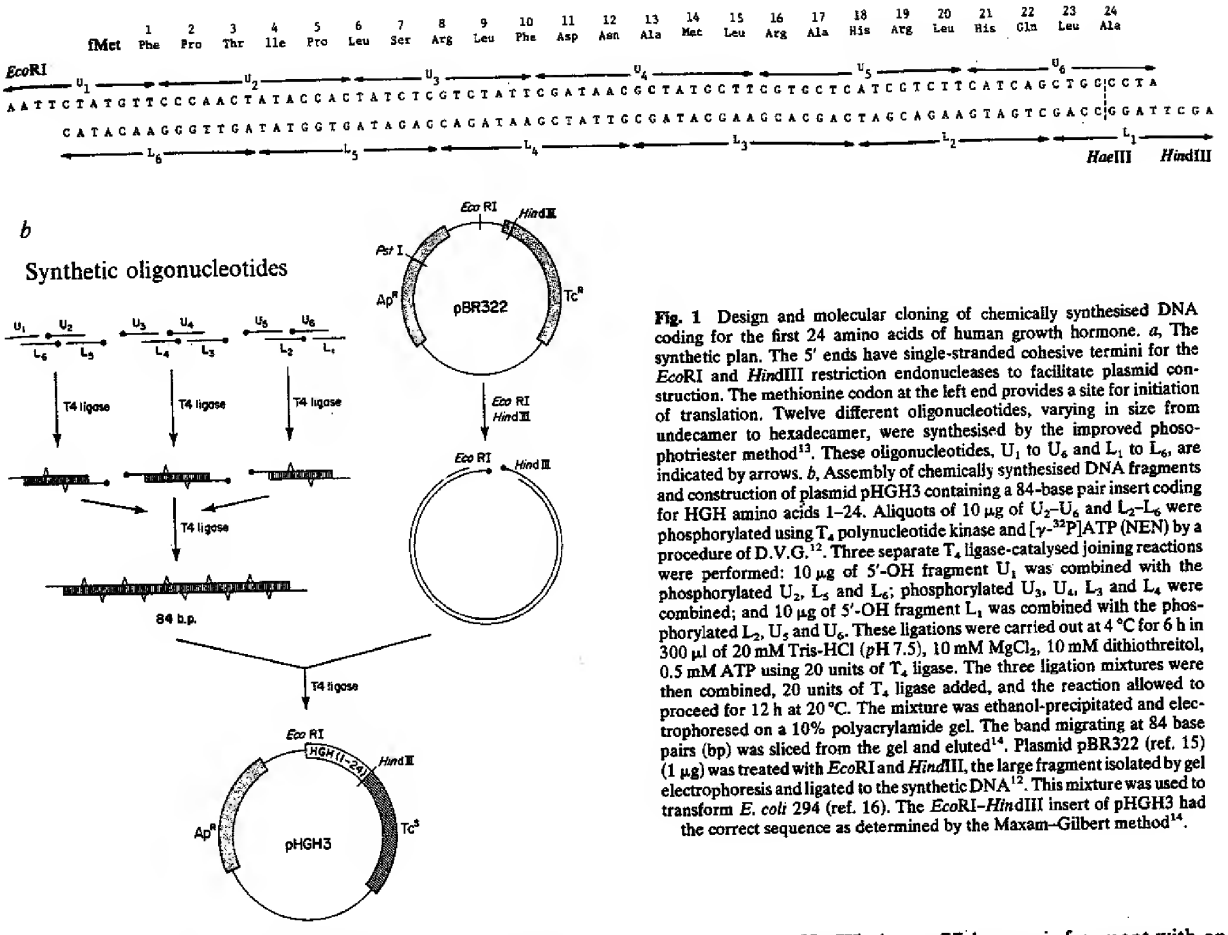


Fig. 1 Design and molecular cloning of chemically synthesised DNA coding for the first 24 amino acids of human growth hormone. *a*, The synthetic plan. The 5′ ends have single-stranded cohesive termini for the *Eco*RI and *Hin*dIII restriction endonucleases to facilitate plasmid construction. The methionine codon at the left end provides a site for initiation of translation. Twelve different oligonucleotides, varying in size from undecamer to hexadecamer, were synthesised by the improved phosphotriester method[13]. These oligonucleotides, U_1 to U_6 and L_1 to L_6, are indicated by arrows. *b*, Assembly of chemically synthesised DNA fragments and construction of plasmid pHGH3 containing a 84-base pair insert coding for HGH amino acids 1–24. Aliquots of 10 μg of U_2–U_6 and L_2–L_6 were phosphorylated using T_4 polynucleotide kinase and [γ-^{32}P]ATP (NEN) by a procedure of D.V.G.[12]. Three separate T_4 ligase-catalysed joining reactions were performed: 10 μg of 5′-OH fragment U_1 was combined with the phosphorylated U_2, L_5 and L_6; phosphorylated U_3, U_4, L_3 and L_4 were combined; and 10 μg of 5′-OH fragment L_1 was combined with the phosphorylated L_2, U_5 and U_6. These ligations were carried out at 4 °C for 6 h in 300 μl of 20 mM Tris-HCl (*p*H 7.5), 10 mM $MgCl_2$, 10 mM dithiothreitol, 0.5 mM ATP using 20 units of T_4 ligase. The three ligation mixtures were then combined, 20 units of T_4 ligase added, and the reaction allowed to proceed for 12 h at 20 °C. The mixture was ethanol-precipitated and electrophoresed on a 10% polyacrylamide gel. The band migrating at 84 base pairs (bp) was sliced from the gel and eluted[14]. Plasmid pBR322 (ref. 15) (1 μg) was treated with *Eco*RI and *Hin*dIII, the large fragment isolated by gel electrophoresis and ligated to the synthetic DNA[12]. This mixture was used to transform *E. coli* 294 (ref. 16). The *Eco*RI–*Hin*dIII insert of pHGH3 had the correct sequence as determined by the Maxam–Gilbert method[14].

sequences by the Grunstein–Hogness hybridisation procedure[21] using ^{32}P-labelled human chorionic somatomammotropin cDNA[22] as probe. Seven colonies hybridised specifically with the probe, which is nearly identical to HGH in nucleotide sequence[5]. Three had identical *Hae*III inserts of approximately 550 base pairs containing *Sma*I, *Bgl*II, *Pst*I and *Pvu*II sites, all characteristic of HGH cDNA[5]. Sequence analysis of the insert from one of these clones, pHGH31 (Fig. 3) gave only one discrepancy (CUG versus CUA codon for leucine 101) when compared with the previously determined cDNA sequence[5]. The amino acid sequence derived from the cDNA sequence differs from the published amino acid sequences of Li[23] at positions 74 (Glu versus Gln), 107 (Asp versus Asn) and 109 (Asn versus Asp). Although we cannot rule out post-translational modification, the discrepancies probably stem from shortcomings of earlier protein sequencing methods in the identification of amidated amino acids.

Construction and cloning of HGH expression plasmids

To express directly HGH in *E. coli*, a plasmid (pGH6) having two *lac* promoters was constructed as follows. A 285-base pair *Eco*RI fragment containing two 95-base pair UV-5 *lac* promoter fragments separated by a 95-base pair heterologous DNA fragment was isolated from the plasmid pKB268 (ref. 24) and inserted into pBR322 at the *Eco*RI site. A plasmid (pGH1) was isolated with the promoter orientated to initiate transcription in the direction of the tetracycline resistance (tc^R) gene. The *Eco*RI site distal to the tc^R gene was destroyed by a previously outlined procedure[11].

Figure 4 reviews the steps involved in assembling the two separately cloned HGH DNA fragments. Treatment of pHGH3 with *Eco*RI and *Hae*III gives a 77-base pair fragment with an *Eco*RI cohesive end and a *Hae*III blunt end (also see Fig. 1*a*). This fragment, which codes for HGH amino acids 1–23, was isolated from a polyacrylamide gel. The plasmid pHGH31 was cleaved with *Hae*III and the 55-base pair HGH cDNA fragment isolated by gel electrophoresis. *Xma*I treatment of this fragment yielded a 512-base pair fragment coding for amino acids 24–191 of HGH with one *Hae*III blunt end and one *Xma*I sticky end. The ligation of these two fragments resulted in the formation of high molecular weight DNA. Treatment with *Sma*I (which recognises the same sequence as *Xma*I but leave flush ends) and *Eco*RI converted the ligation products to three distinct fragments: a 154-base pair dimer of the 77-base pair DNA fragment, a 1,024-base pair dimer of the 512-base pair fragment, and the desired 591-base pair DNA coding for the entire HGH sequence. This latter fragment was purified by gel electrophoresis and inserted into pGH6 between the *Eco*RI and the S_1 nuclease-treated *Hin*dIII sites. Following transformation of χ1776, the expression plasmid pHGH107 was identified by colony screening and restriction analysis as described in Fig. 3. The HGH DNA sequence of pHGH107 was verified by the Maxam–Gilbert procedure[14].

In our plan to express HGH two base pairs were included between the *Eco*RI AATT 'sticky' end and the ATG codon of the chemically synthesised portion of the HGH coding sequence (see Fig. 1). Therefore, in the plasmid pHGH107, 11 base pairs separate the *lac* AGGA ribosome binding site[26] and the ATG translational start for HGH:

5′...AGGAAACAG|AATTCTATG...
3′...TCCTTTGTCTTAA|GATAC...

In the naturally occurring *lac* system the separation between

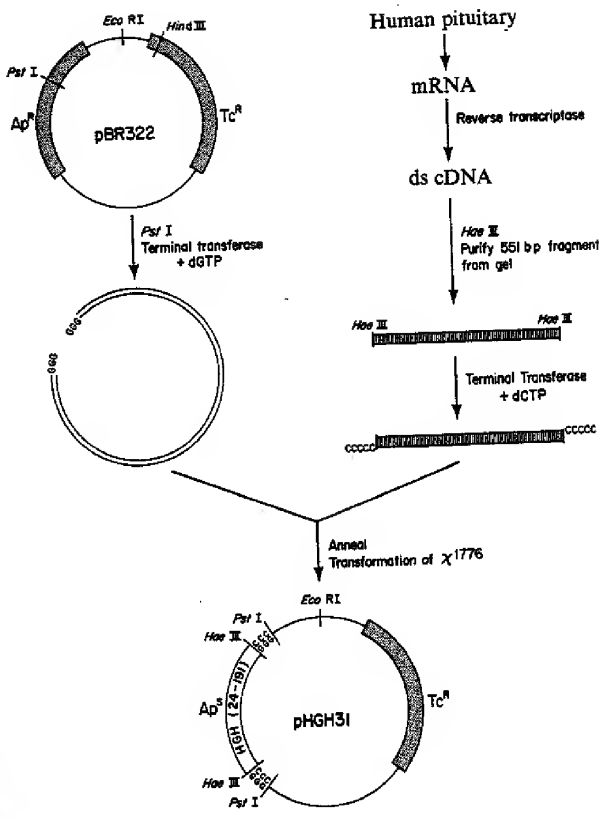


Fig. 2 Construction of a plasmid (pHGH31) containing coding sequences for amino acids 24–191 of human growth hormone. Poly(A)-mRNA was prepared from human pituitaries by a published procedure[17]. ds-cDNA (1.5 μg) was prepared from 5 μg of this RNA essentially as described by Wickens *et al.*[18], except that DNA polymerase 'Klenow fragment'[19] was substituted for DNA polymerase I in the second strand synthesis. cDNA (90 ng) was treated with *Hae*III, electrophoresed on an 8% polyacrylamide gel, and the region around 550 base pairs eluted. Approximately 1 ng of cDNA was obtained. Terminal deoxynucleotidyl transferase (TdT) was used to add approximately 20 dC residues per 3′ terminus[10]. Annealing of the dC-tailed ds-cDNA with the dG-tailed vector DNA (60 ng) was performed in 130 μl of 10 mM Tris-HCl (*p*H 7.5), 100 mM NaCl, 0.25 mM EDTA. The mixture was heated to 70 °C, allowed to cool slowly to 37 °C (12 h), then to 20 °C (6 h), before being used to transform χ1776 (ref. 20) by a published procedure[4].

ribosome binding site and initiation codon is 7 base pairs. The exact *lac* sequence in this region was restored as follows. The plasmid was cleaved with *Eco*RI, treated with S_1 nuclease to digest the single stranded ends, and reclosed by blunt-ended ligation. The plasmids from three individual transformants of χ1776 were analysed. In all instances the *Eco*RI site had been removed and a new *Alu*I site created in the plasmid (pHGH107–1).

5′...AGGAAACAG↓CTATG...
3′...TCCTTTGTC↓GATAC...

This structure, in which the ATG codon is now seven nucleotides downstream from the ribosome binding site, might be expected to increase the efficiency with which HGH mRNA is translated.

HGH radioimmune activity in bacterial extracts

Extracts of *E. coli* χ1776 containing the expression plasmids were tested for HGH expression by a direct radioimmunoassay (RIA) (Table 1). Easily detectable amounts of HGH were produced by bacteria containing pHGH107 or the derivative of pHGH107 lacking the *Eco*RI site (pHGH107–1). Contrary to

```
                                                                    40
      24
      Ala Phe Asp Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro Gln Thr Ser
5'...G GCC UUU GAC ACC UAC CAG GAG UUU GAA GAA GCC UAU AUC CCA AAG GAA CAG AAG UAU UCA UUC CUG CAG AAC CCC CAG ACC UCC
                                                                        Pst I
                                                                                                              80
                                    60
Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser Leu Leu
CUC UGU UUC UCA GAG UCU AUU CCG ACA CCC UCC AAC AGG GAG GAA ACA CAA CAG AAA UCC AAC CUA GAG CUG CUC CGC AUC UCC CUG CUG

                                                                   100
Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr
CUC AUC CAG UCG UGG CUG GAG CCC GUG CAG UUC CUC AGG AGU GUC UUC GCC AAC AGC CUA GUG UAC GGC GCC UCU GAC AGC AAC GUC UAU

                                                                                                             140
                                   120
Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln
GAC CUC CUA AAG GAC CUA GAG GAA GGC AUC CAA ACG CUG AUG GGG AGG CUG GAA GAU GGC AGC CCC CGG ACU GGG CAG AUC UUC AAG CAG
                                                                                                  Bgl II
                                                                   160
Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp
ACC UAC AGC AAG UUC GAC ACA AAC UCA CAC AAC GAU GAC GCA CUA CUC AAG AAC UAC GGG CUG CUC UAC UGC UUC AGG AAG GAC AUG GAC

                                   180                                                                      191
Lys Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe Stop
AAG GUC GAG ACA UUC CUG CGC AUC GUG CAG UGC CGC UCU GUG GAG GGC AGC UGU CGC UUC UAG CUGCCCGGGUGGCAUCCCUGUGACCCCUCCCCAGU
                                                  Pvu II                          Sma I
                                                                                  Xma I

GCCUCUCCUGGCC....3'
      Bae III
```

Fig. 3 The amino acid and mRNA sequences of HGH as determined by DNA sequencing[14] of pHGH31. The nucleotide sequence from the *Sma*I site to the *Hae*III site in the 3′ noncoding region is from ref. 5, and was not determined here.

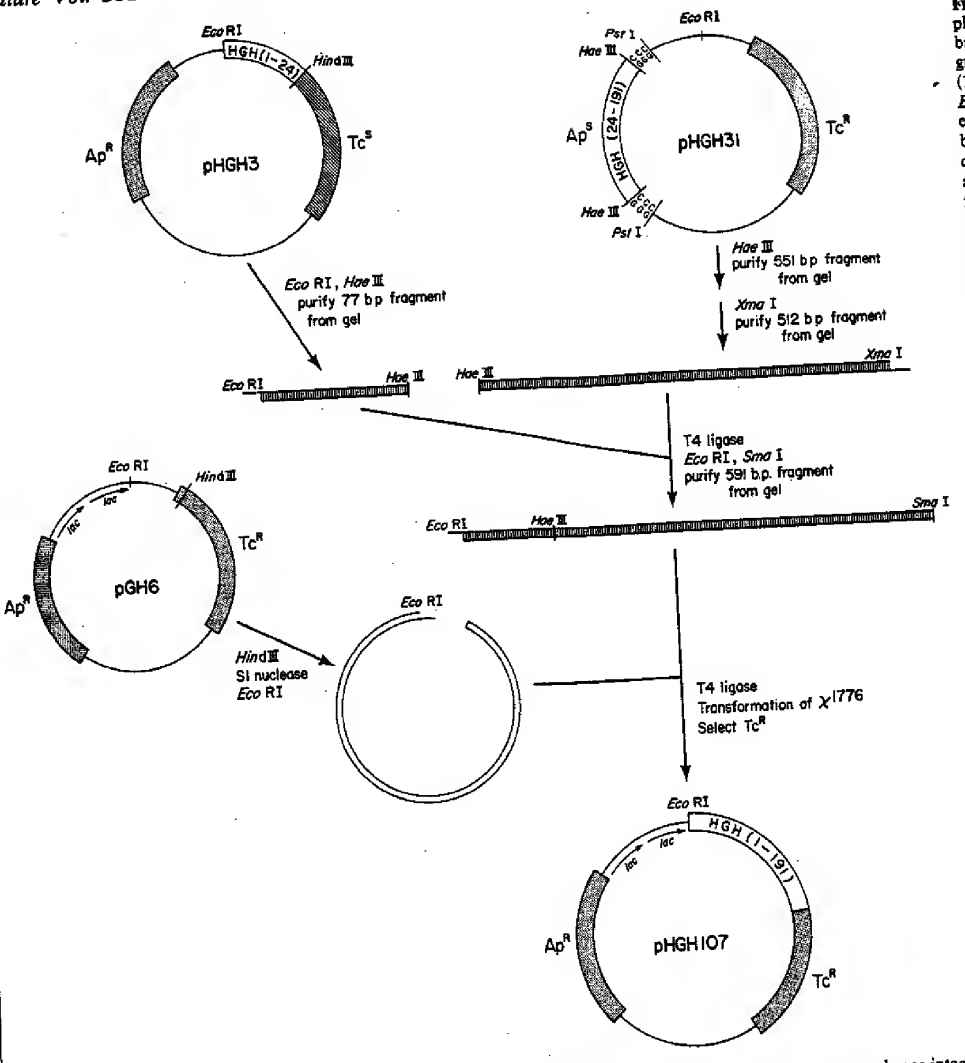


Fig. 4 Construction of a plasmid (pHGH107) for the bacterial expression of human growth hormone. pHGH3 (10 μg) was cleaved with *Eco*RI and *Hae*III restriction endonucleases and the 77-base pair fragment containing coding sequences for HGH amino acids 1–23 was isolated from an 8% polyacrylamide gel. The plasmid pHGH31 (~5 μg) was cleaved with *Hae*III. The 551-base pair HGH sequence and a co-migrating 540-base pair *Hae*III fragment of pBR322 were purified by gel electrophoresis. Subsequent treatment with *Xma*I cleaved only the HGH sequence, removing 39 base pairs from the 3′ noncoding region. The resulting 512-base pair fragment was purified from the 540-base pair pBR322 *Hae*III piece by electrophoresis on a 6% polyacrylamide gel. Samples of 0.3 μg of the 77-base pair *Eco*RI–*Hae*III fragment and 0.3 μg of the 512-base pair *Hae*III–*Xma*I fragment were polymerised with T_4 DNA ligase in a 16-μl reaction for 14 h at 4 °C. The mixture was heated at 70 °C for 5 min to inactivate the ligase, then treated with *Eco*RI (to cleave fragments which had dimerised through their *Eco*RI sites) and with *Sma*I (to cleave *Xma*I dimers), yielding a 591-base pair fragment with an *Eco*RI 'sticky' end and a *Sma*I 'blunt' end. After purification on a 6% polyacrylamide gel, approximately 30 ng of this fragment were obtained. The expression plasmid pGH6 containing tandem *lac* UV-5 promoters, was treated successively with *Hin*dIII, nuclease S_1, and *Eco*RI and purified by gel electrophoresis. This removes most of the promoter sequence of the tc^R gene, but leaves the structural gene intact (25). 50 ng of the resulting vector, which had one *Eco*RI sticky end and one blunt end, was ligated to 10 ng of the 591-base pair HGH DNA. The ligation mixture was used to transform χ1776. Colonies were selected for growth on tetracycline (12.5 g ml^{-1}). Since the tc^R promoter is no longer functional, tetracycline resistance is dependent on transcription from the *lac* promoters reading through the HGH gene sequence into the tc^R gene. Approximately 400 transformants were obtained. Colony screening by filter hybridisation[22] identified 12 colonies containing HGH sequences. The plasmids isolated from three of these colonies gave the restriction patterns expected for the correctly assembled gene when cleaved with *Hae*III, *Pou*II and *Psi*. The DNA sequence of one clone, pHGH107, was determined.

our expectations, the introduction of four extra base pairs of DNA into the *lac* sequence in the region between the ribosome binding site and initiation codon results in increased HGH synthesis.

To verify that HGH expression is under *lac* operon control the plasmid pHGH107 was transformed into the *E. coli* K12 strain D1210, an i^Q derivative of *E. coli* K12 HB101 which overproduces *lac* repressor[12]. The data in Table 1 demonstrate that repressor effectively blocks HGH production and that this effect can be reversed by the addition of the inducer isopropyl-β-D-thiogalactoside.

Identification of HGH from bacterial extracts

Extracts of *E. coli* K12 strain RV308 (ref. 27) containing either pHGH107 or pBR322 were electrophoresed on SDS-polyacrylamide gels. The resulting protein patterns show one clear difference (Fig. 5*a*)—an additional protein, which co-migrates with HGH isolated from pituitaries, appears in the RV308/pHGH107 extract. To achieve partial purification of HGH, an extract of a 1-litre culture of χ1776/pHGH107 was prepared and fractionated as described in Fig. 5 legend. The polyacrylamide gel electrophoresis pattern of this partially

Table 1 Radioimmunoassay of HGH in extracts from bacteria containing expression plasmids

Strain	Cell density (cells per ml)	HGH by RIA (μg per ml)	HGH copies per cell
χ1776/pHGH107	3.69×10^8	2.4	186,000
	1×10^9	1.4	39,000
χ1776/pHGH107-1	3.6×10^8	1.5	116,000
	1×10^9	0.5	14,000
χ1776/pBR322	3.6×10^8	0	0
D1210/pHGH107	3.8×10^8	2×10^{-4}	15
D1210/pHGH107 (IPTG)	3.8×10^8	1.0	75,000

E. coli strains containing the appropriate plasmid were grown to the indicated cell density and collected by centrifugation. Stationary phase cultures of χ1776 contain 10^9 cells per ml. The cell pellet was resuspended in 55 μl of 50 mM Tris-HCl (pH 8), 50 mM EDTA, 15% sucrose, 1 mg ml^{-1} lysozyme, 0.025% lithium dodecylsulphate. After 30 min at 0 °C, 10 μl of 150 mM Tris-HCl (pH 7.5), 280 mM $MgCl_2$, 4 mM $CaCl_2$ and 1 μg of DNase I were added. The mixture was centrifuged for 15 min at 12,000 g. Serial dilutions of the supernatants were analysed by radioimmunoassay using the Phadebas HGH PRIST kit (Pharmacia). Values are the average of three independent experiments expressed as μg per ml of cell culture. With HGH standard, the detection limit in 1 ml of culture was 0.02 ng. No activity was found in extracts of χ1776/pBR322. Where indicated, isopropyl-β-D-thiogalactoside (IPTG) was added to the cell culture at a concentration of 2 mM.

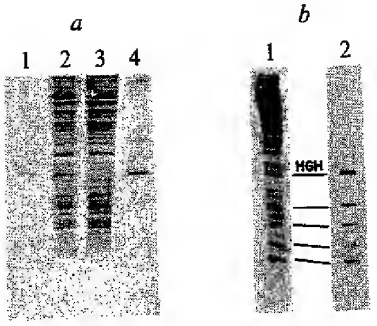


Fig. 5 Identification of HGH produced in bacteria by SDS-polyacrylamide gel electrophoresis. *a*, Protein patterns of crude extracts and partially purified HGH stained with Coomassie brilliant blue. Slot 1 contains 0.5 μg of pituitary HGH standard (Kabi), slot 2 contains a cell lysate of RV308/pHGH107, slot 3 contains a cell lysate of RV308/pBR322, slot 4 contains partially purified HGH isolated from χ1776/pHGH107. The samples were separated on a 15% polyacrylamide slab gel using the buffer system of Maizel[28] with the addition of 6 M urea. Crude lysates were prepared by growing cells in LB with 5 μg ml^{-1} tetracycline followed by lysis in 2% SDS, 1% β-mercaptoethanol. The lysates were precipitated with 10 volumes of cold acetone and the pellets were redissolved in SDS sample buffer for use in gel electrophoresis. The partially purified HGH was prepared from a stationary phase culture of χ1776/pHGH107. Cells were collected and resuspended in 1/50 of their original volume in 30 mM potassium phosphate (*p*H 7.0) containing 0.05 M NaCl and lysed by sonication. Polyethylenimine (Miles, Polymin-P) was added to 0.2%. After centrifugation for 1 h at 100,000*g*, ammonium sulphate was added to the supernatant to 60% saturation. The ammonium sulphate pellet was dissolved in 2 ml 10 mM potassium phosphate (*p*H 7.0), 0.5 M NaCl and chromatographed on a Sephacryl S-200 column (2.0 × 50 cm) equilibrated in the same buffer. The radioimmune active peak (assayed with a Pharmacia Phadebus HGH kit) was pooled, the protein concentrated by ammonium sulphate precipitation, the pellet redissolved in 1/10 volume of buffer, and the solution dialysed against 10 mM potassium phosphate (*p*H 7.0), 0.5 M NaCl. Precipitated material was removed by centrifugation resulting in a HGH preparation of approximately 25% purity. *b*, Autoradiograms of ^{35}S-labelled extracts of RV308/pHGH107. Slot 1 contains a total lysate of RV308/pHGH107 labelled with $H_2{}^{35}SO_4$. Slot 2 contains an ^{35}S-labelled extract of RV308/pHGH107 precipitated with α-HGH antiserum. The major (top) band co-migrates with unlabelled HGH standard (not shown). Cultures (1 ml) of RV308/pHGH107 were grown to $A_{550} = 1$ in low sulphur medium[29] containing 0.2 mCi ml^{-1} $H_2{}^{35}SO_4$, chased with 10 mM $MgSO_4$ for 5 min, collected and lysed using Triton X-100 and lysozyme[30]. Following DNase and RNase treatment[29], the lysate was mixed with a 10-fold excess of unlabelled RV308/pBR322 extract and diluted 1:1 into Triton immunoprecipitation buffer (0.15 M NaCl, 1% Triton X-100, 0.05 M Tris-HCl *p*H 7.5). Twenty μl of α-HGH antiserum (Kabi) was added per ml of original culture and the reaction was incubated for 12 h at 4 °C. The mixture was centrifuged and the supernatant incubated for 2 h with formaldehyde fixed *Staphylococcus* cells[31], filtered on 0.45-μm Nucleopore polycarbonate filters, washed with Triton immunoprecipitation buffer and extracted with SDS sample buffer[28]. The samples were run on a 15% slab gel containing urea and SDS as described above.

purified sample is shown in Fig. 5*a*, slot 4. The most prominent band co-migrates with HGH standard and its relative intensity is consistent with the 25% purity calculated from radioimmune activity.

In an independent attempt to confirm that the antigenically active component in the bacterial extracts co-migrates with HGH in SDS gels, we precipitated extracts from cells labelled with ^{35}S with antiserum prepared against HGH and separated the precipitated proteins by polyacrylamide gel electrophoresis. The autoradiogram of the immunoprecipitated material (Fig. 5*b*, slot 2) shows a major band, with electrophoretic mobility identical to that of HGH, as well as four smaller bands. All five of these bands are present in the unfractionated ^{35}S-labelled lysate (slot 1), whereas only the band with the same mobility as HGH can be seen in unlabelled extracts of cells grown in LB broth (Fig. 5*a*, slot 2). All five immune precipitated bands can be competed out by excess cold HGH (data not shown) and are thus antigenically related. This suggests that the four smaller HGH-related products observed in minimal media may be due to proteolytic degradation.

Additional evidence of proteolytic degradation of HGH is presented in Table 1. The levels of HGH in extracts from strains containing pHGH107 or pHGH107-1 were considerably higher in log-phase cultures than in stationary-phase cultures. Strains transformed with several other plasmids whose products are under *lac* promoter control show no comparable decrease of gene expression as a function of growth phase (unpublished results from this laboratory). It thus seems that there is some turnover of HGH in bacterial cells although we cannot rule out other explanations for the four smaller HGH-related products in the ^{35}S-labelled extracts such as premature termination of translation.

Despite the observed instability of HGH in certain growth conditions, the molecule seems to be relatively stable in log-phase cultures grown in rich media. In these conditions the amount of HGH in the cells (186,000 monomers per cell, Table 1) compares favourably with the expression level of other cloned genes using the same promoter in optimised conditions[32].

Conclusion

Using a novel combination of chemically synthesised DNA and cDNA, a recombinant *E. coli* strain has been constructed which produces HGH in large amounts. This is the first time that a human polypeptide has been directly expressed in *E. coli* in a non-precursor form. The hybrid DNA cloning techniques described as a route to the cloning and expression of HGH coding sequences in *E. coli* are generally applicable to other polypeptides which are synthesised initially as inactive precursors and later processed, or for which full length cDNA transcripts are unavailable.

We thank John Shine for help in planning the project, Elizabeth McLeod for assistance, Ron Wetzel for help in peptide identification, Frances Fields and Dale Cullins for technical support, Alane Gray for preparing illustrations, and Robert Swanson and Herbert Boyer for their encouragement. RNA used was prepared by P.H.S. while a postdoctoral fellow in the Department of Biochemistry and Biophysics, University of California at San Francisco. Additional RNA was obtained from pituitaries donated by Drs Gerald Silverberg and O. H. Pearson. This work was supported by a contract from KabiGen AB. All experiments involving recombinant genomes were performed in accordance with current NIH guidelines.

Received 6 July; accepted 16 August 1979.

1. Raben, M. S. *J. clin. Endocr.* **18**, 901–904 (1958).
2. *Proc. of NIAMDD Symp.* Publ. No. 74–612 (ed. Raiti, S.) (Baltimore, Maryland, 1973).
3. Bancroft, F. C. *Expl Cell Res.* **79**, 275–278 (1973).
4. Seeburg, P. H., Shine, J., Martial, J. A., Baxter, J. D. & Goodman, H. M. *Nature* **270**, 486–494 (1977).
5. Goodman, H. M. *et al.* in *Specific Eukaryotic Genes* (eds Engberg, J., Klenow, H. & Leick, V.) 179–190 (Munskagaard, Copenhagen, 1979).
6. Villa-Komaroff, L. *et al. Proc. nam. Acad. Sci. U.S.A.* **75**, 3727–3731 (1978).
7. Seeburg, P. H. *et al. Nature* **276**, 795–798 (1978).
8. Mercereau-Puijalon, O. *et al. Nature* **275**, 505–510 (1978).
9. Fraser, T. H. & Bruce, B. J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5936–5940 (1978).
10. Chang, A. C. Y. *et al. Nature* **275**, 617–624 (1978).
11. Itakura, K. *et al. Science* **198**, 1056–1063 (1977).
12. Goeddel, D. V. *et al. Proc. nam. Acad. Sci. U.S.A.* **76**, 106–110 (1979).
13. Crea, R., Kraszewski, A., Hirose, T. & Itakura, K. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5765–5769 (1978).
14. Maxam, A. M. & Gilbert, W. *Proc. nam. Acad. Sci. U.S.A.* **74**, 560–564 (1978).
15. Bolivar, F. *et al. Gene* **2**, 95–113 (1977).
16. Backman, K., Ptashne, M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **73**, 4174–4178 (1976).
17. Ulrich, A. *et al. Science* **196**, 1313–1319 (1977).
18. Wickens, M. P., Buell, G. N. & Schimke, R. T. *J. biol. Chem.* **253**, 2483–2495 (1978).
19. Klenow, H. & Henningsen, I. *Proc. natn. Acad. Sci. U.S.A.* **65**, 168–171 (1970).
20. Curtiss, R. III *et al.* in *Molecular Cloning of Recombinant DNA* (eds Scott, W. A. & Werner, R.) 99–111 (Academic, New York, 1977).
21. Grunstein, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961–3965 (1975).
22. Shine, J., Seeburg, P. H., Martial, J. A., Baxter, J. D. & Goodman, H. M. *Nature* **270**, 494–499 (1977).
23. Li, C. H. *Proc. Am. phil. Soc.* **116**, 365 (1972).
24. Backman, K. & Ptashne, M. *Cell* **13**, 65–71 (1978).
25. Boyer, H. W. *et al.* in *Recombinant Molecules: Impact on Science and Society*, (eds Beers, R. F. & Bassett, E. G.) 9–20 (Raven, New York, 1977).
26. Maizels, N. *Nature* **249**, 647 (1974).
27. Mauer, R. thesis, Harvard Univ. (1978).
28. Maizel, J. V. Jr *Meth. Virol.* **5**, 180–246 (1971).
29. Roberts, J. & Roberts, C. *Proc. nam. Acad. Sci. U.S.A.* **72**, 147–151 (1975).
30. Katz, L., Kingsbury, D. T. & Helinski, D. R. *J. Bact.* **144**, 577–591 (1973).
31. Kessler, S. W. *J. Immun.* **177**, 1482–1490 (1976).
32. Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K. & Riggs, A. D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 106–110 (1979).

TAB V

US005460811A

United States Patent [19]

Goeddel et al.

[11] Patent Number: 5,460,811

[45] Date of Patent: Oct. 24, 1995

[54] **MATURE HUMAN FIBROBLAST INTERFERON**

[75] Inventors: **David V. Goeddel**; **Roberto Crea**, Burlingame, Calif.

[73] Assignee: **Genentech, Inc.**, South San Francisco, Calif.

[21] Appl. No.: **365,284**

[22] Filed: **Jun. 12, 1989**

Related U.S. Application Data

[60] Continuation of Ser. No. 889,722, Jul. 28, 1986, abandoned, which is a division of Ser. No. 291,892, Aug. 11, 1981, abandoned, which is a continuation-in-part of Ser. No. 190,799, Sep. 25, 1980, abandoned.

[51] **Int. Cl.**[6] **C07K 15/26**; A61K 37/66

[52] **U.S. Cl.** **424/85.6**; 424/85.4; 530/351

[58] **Field of Search** 530/351; 424/85.6, 42485.4

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,262,090	4/1981	Colby et al.	435/97
4,289,689	9/1981	Friesen et al.	260/112 R
4,332,892	6/1982	Ptashne et al.	435/172
4,342,832	8/1982	Goeddel et al.	435/172
4,418,149	4/1983	Ptashne et al.	435/172
4,874,702	10/1989	Fiers et al.	435/172
5,015,730	5/1991	Friesen et al.	530/351

FOREIGN PATENT DOCUMENTS

6694	1/1980	European Pat. Off. .
2007676	5/1979	United Kingdom .

OTHER PUBLICATIONS

Darnell et al. (Darnell), Molecular Cell Biology, "Golgi Vesicles: Sorting and Glycosylation of Secretory and Membrane Proteins", Second Edition, pp. 661–667 (1990).

Houghton et al., Nature 285:536 (published Jun. 19, 1980).

Goeddel et al., Nature 287:411–416 (published Oct. 2, 1980).

Horoszewicz et al., Cancer Treatment Rep. 62:1899–1906 (1978)

Stewart II et al., Virology 97:473–476 (1979).

Pestka et al., Proc. Nat'l Acad. Sci USA 72: 3898–3901 (1975).

Reynolds et al., Proc. Nat'l Acad. Sci. USA 72: 4881–4885 (1975).

Cavalieri et al., Proc. Nat'l Acad. Sci. USA 74: 4415–4419 (1977).

Stewart II, in The Interferon System (Springer–Verlag, New York, 1979), pp. 90–96, 335, 386 and 389.

Grob et al., Biochemistry 18: 5782–86 (1979).

Havell et al., Virology 63: 475–83 (1975).

Houghton et al., Nucleic Acids Res. 8: 1913–31 (1980).

Bose et al., J. Biol. Chem. 251: 1659–62 (1976).

Davey et al., Biochemistry 15: 704–13 (1976).

Eagon et al., J. Biol. Chem. 252: 2372–83 (1977).

Mizrahi et al., J. Biol. Chem. 253: 7612–15 (1978).

Stewart et al., Proc. Natl. Acad. Sci. USA 74: 4200–04 (1977).

Struck et al., In *The Biochemistry of Glycoproteins and Proteoglycans*, Plenum Press (W. J. Lennarz ed.), 1980, pp. 35–83.

Taniguchi et al., Proc. Japan Acad. 55(Ser.B):464–469 (1979).

Taniguchi et al., Gene 10:11–15 (published May 19, 1980).

Taniguchi et al., Proc. Nat'l. Acad. Sci. USA 77:4003–4006 (published Aug. 13, 1980).

Taniguchi et al., Proc. Nat'l Acad. Sci. USA 77:5230–5233 (published Oct. 15, 1980).

Taniguchi et al., Nature 285:547–549 (published Jun. 19, 1980).

Derynck et al., Nature 285:542–547 (published Jun. 19, 1980).

Derynck et al., Nature 287:193–197 (published Sep. 18, 1980).

Guarente et al., Science 209:1428–1430 (published Sep. 19, 1980).

Anonymous, Research Disclosure (Jul. 1979).

Houghton et al., Nucleic Acids Res. 8(13):2885–2894 (published Jul. 14, 1980).

Goeddel et al., Nucleic Acids Res. 8(18):4057–4074 (published Sep. 25, 1980).

Goeddel et al., Nature 281:544–548 (1979)

Chang et al., Nature 275:617–624 (1978).

Backman et al., Cell 13:65–71 (1978).
Knight, Proc. Nat'l Acad. Sci. USA 73(21):520–523 (1976).
Tan et al., J. Biol. Chem. 244:8067–8073 (1979).
Berthold et al., J. Biol. Chem. 253(14):5206–5212 (1978).
Knight et al., Science 207:525–527 (1980).
Reynolds et al., Biochem. Biophys. Res. Comm. 63(1)107–112 (1977).
Havell et al., J. Biol. Chem. 252(12):4425–4427 (1977).
Cavalieri et al., Proc. Nat'l. Acad. Sci. USA 74:3287–3291 (1977).
Pestka et al., Ann. Rev. New York Acad. Sci. 284:697–702 (1977).
Content et al., Virology 122:466–470 (1982).
Roberts et al., Proc. Nat'l. Acad. Sci. USA 76:760–764 (1979).
Roberts et al., Proc. Nat'l. Acad. Sci. USA 76:5596–5600 (1979).
Guarente et al., Cell, 20:543–553 (1980).
Backman et al. Cell 13 pp. 65–71 (1978).
Taniguchi et al. Gene 10 pp. 11–15 (1980).
Taniguchi et al. Proc. Japan Acad. 55 (Ser.B) (1979) pp. 464–469.
Goeddel et al. Nature 281 pp. 544–548 (1979).
Chang et al. Nature 275 pp. 617–624 (1978).
Derynck et al. Nature 285 pp. 542–547 (1980).

Primary Examiner—Garnette D. Draper
Assistant Examiner—Shelly Guest Cermak

[57] **ABSTRACT**

A cDNA library is constructed using mRNA from human fibroblasts induced with poly(I):poly(C). A bacterial clone containing fibroblast interferon cDNA sequences identified by hybridization to a cDNA probe synthesized using deoxyoligonucleotide primers which hybridize to fibroblast interferon mRNA specifically. Expression plasmids are constructed which permit the synthesis in *E. coli* of 8×10^7 units of human fibroblast interferon per liter of culture. The bacterially produced fibroblast interferon is indistinguishable from authentic human fibroblast interferon by several critieria.

6 Claims, 6 Drawing Sheets

Protein	1 2 3 4 Met-Ser-Tyr-Asn-	
mRNA	(5') AUG-UC$^{G}_{A}$$_{U}$$_{C}$-UA$^{U}_{C}$-AA$^{U}_{C}$	(16 combinations)
	(5') AUG-AG$^{U}_{C}$-UA$^{U}_{C}$-AA$^{C}_{U}$	(8 combinations)
Complementary DNA primers	ATT-$^{A}_{G}$TA-$^{T}_{C}$GA-CAT	Pool 1
	ATT-$^{A}_{G}$TA-$^{A}_{G}$GA-CAT	Pool 2
	ATT-$^{A}_{G}$TA-$^{A}_{G}$CT-CAT	Pool 3
	GTT-$^{A}_{G}$TA-$^{T}_{C}$GA-CAT	Pool 4
	GTT-$^{A}_{G}$TA-$^{A}_{G}$GA-CAT	Pool 5
	GTT-$^{A}_{G}$TA-$^{A}_{G}$CT-CAT	Pool 6

FIG. 1

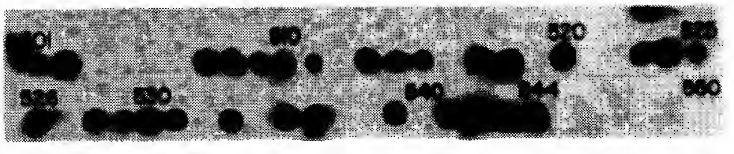


FIG. 2A

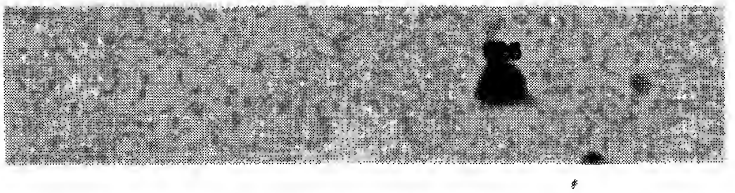


FIG. 2B

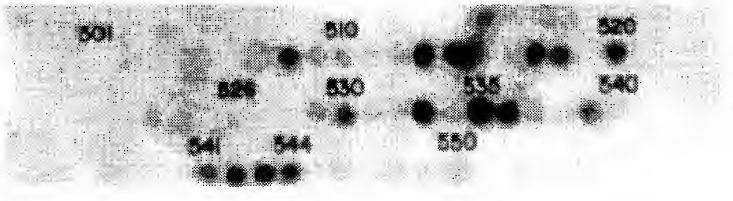


FIG. 2C

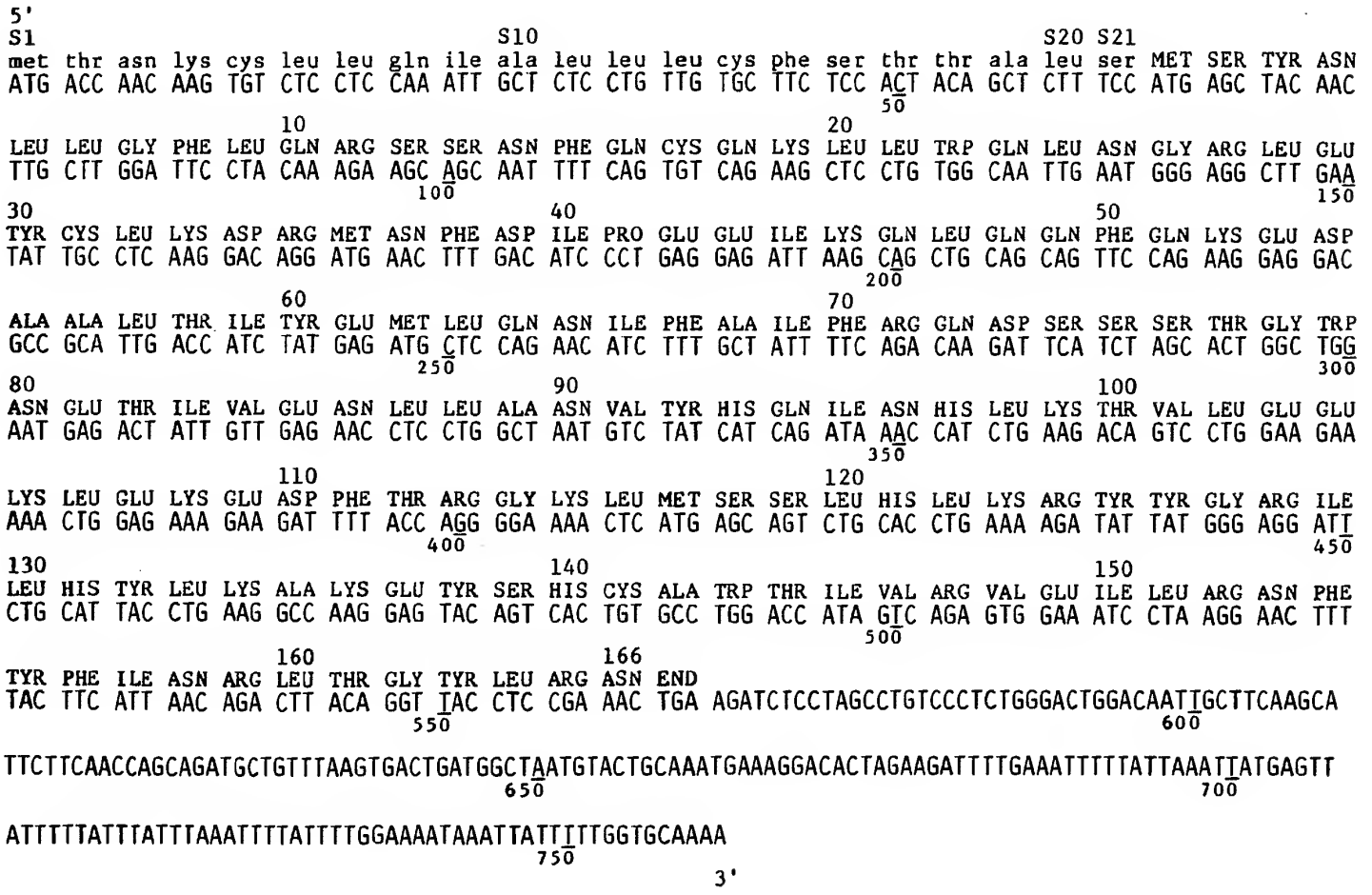


FIG. 3

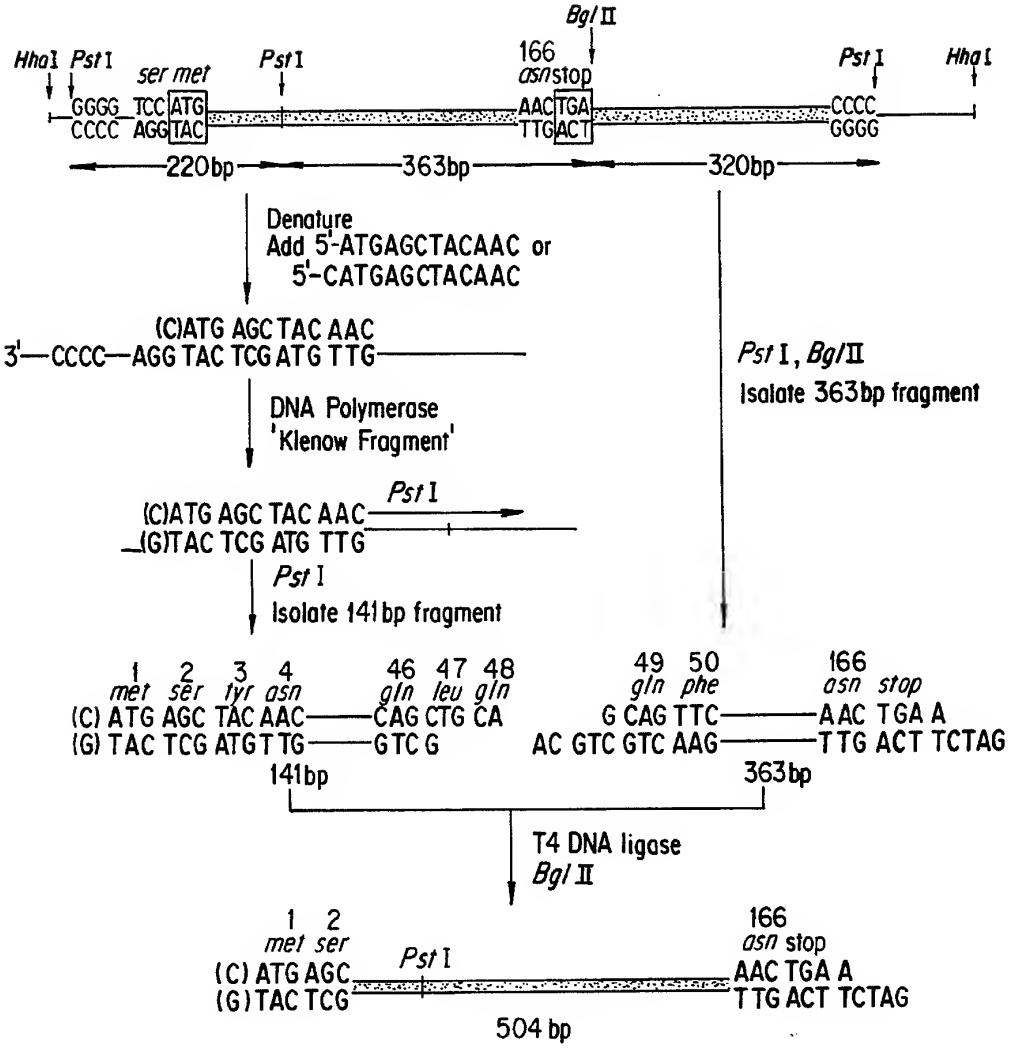

Bgl II
Hha I
Pst I
ser met
Pst I
166
asn stop
Pst I
Hha I
GGGG TCC ATG
CCCC AGG TAC
AAC TGA
TTG ACT
CCCC
GGGG
220bp
363bp
320bp
Denature
Add 5'-ATGAGCTACAAC or
5'-CATGAGCTACAAC
(C)ATG AGC TAC AAC
3'—CCCC—AGG TAC TCG ATG TTG
Pst I, Bgl II
Isolate 363bp fragment
DNA Polymerase
'Klenow Fragment'
Pst I
(C)ATG AGC TAC AAC
(G)TAC TCG ATG TTG
Pst I
Isolate 141bp fragment
1 2 3 4
met ser tyr asn
46 47 48
gln leu gln
(C) ATG AGC TAC AAC—CAG CTG CA
(G) TAC TCG ATG TTG—GTC G
141bp
49 50
gln phe
166
asn stop
G CAG TTC—AAC TGA A
AC GTC GTC AAG—TTG ACT TCTAG
363bp
T4 DNA ligase
Bgl II
1 2
met ser
Pst I
166
asn stop
(C) ATG AGC
(G) TAC TCG
AAC TGA A
TTG ACT TCTAG
504 bp

FIG.4

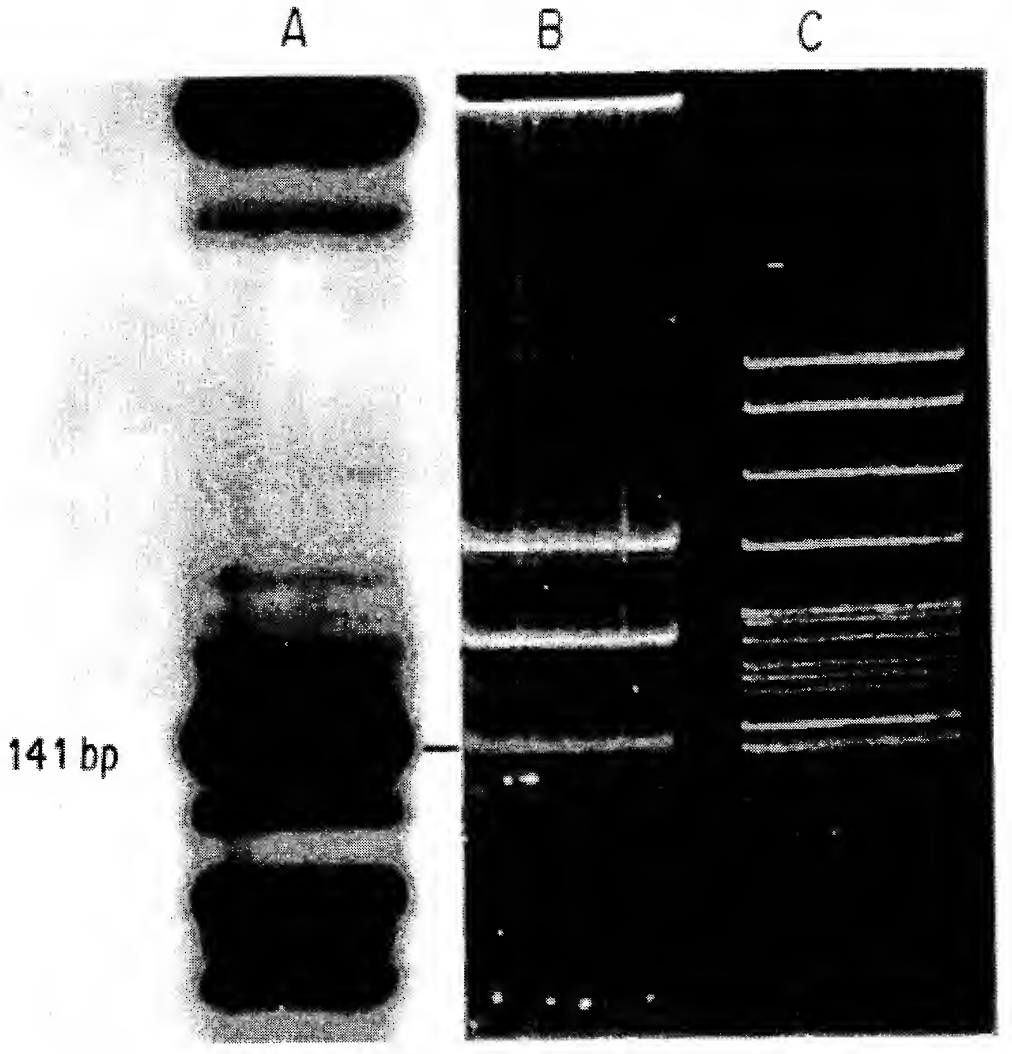


FIG. 5

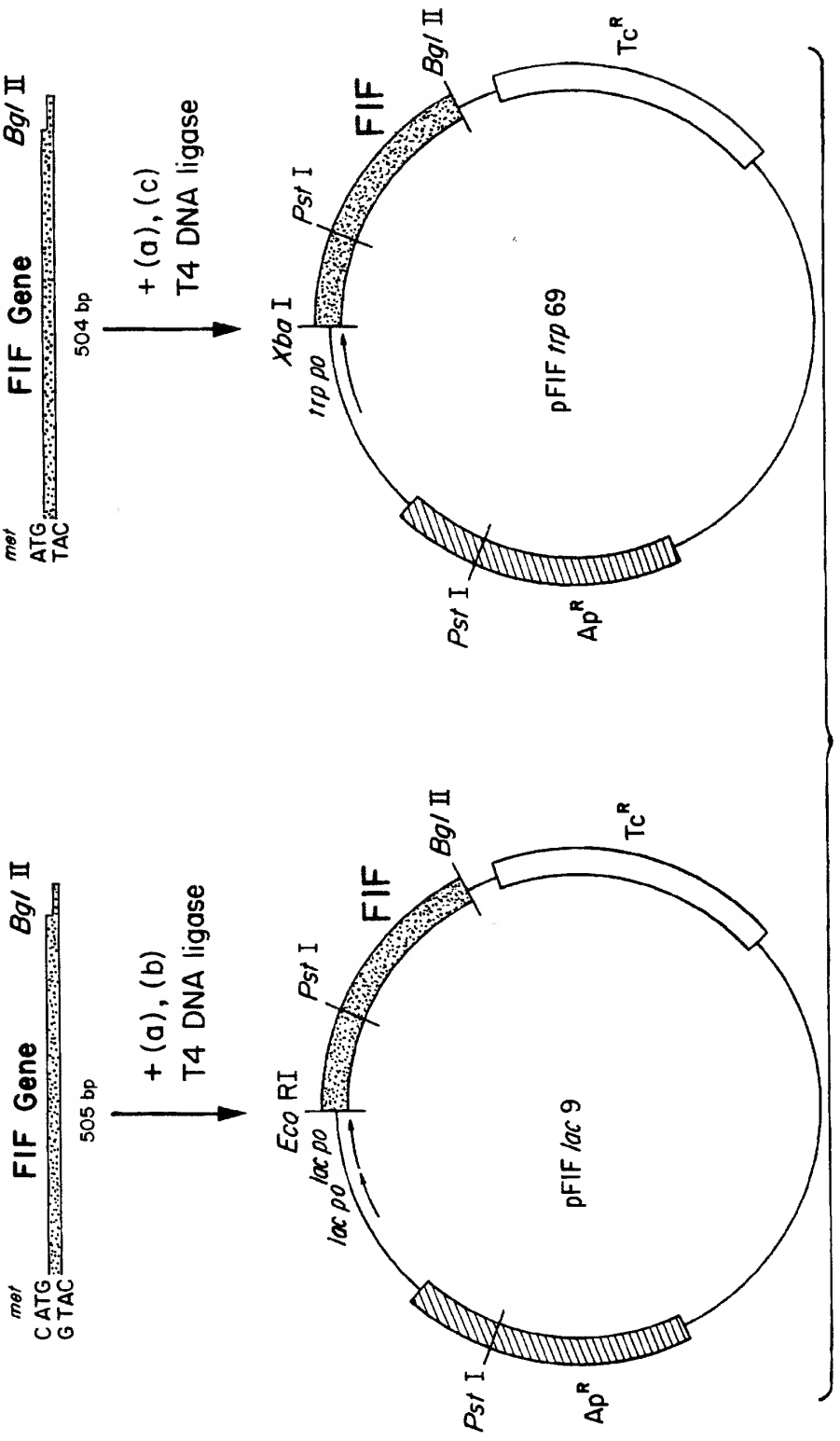
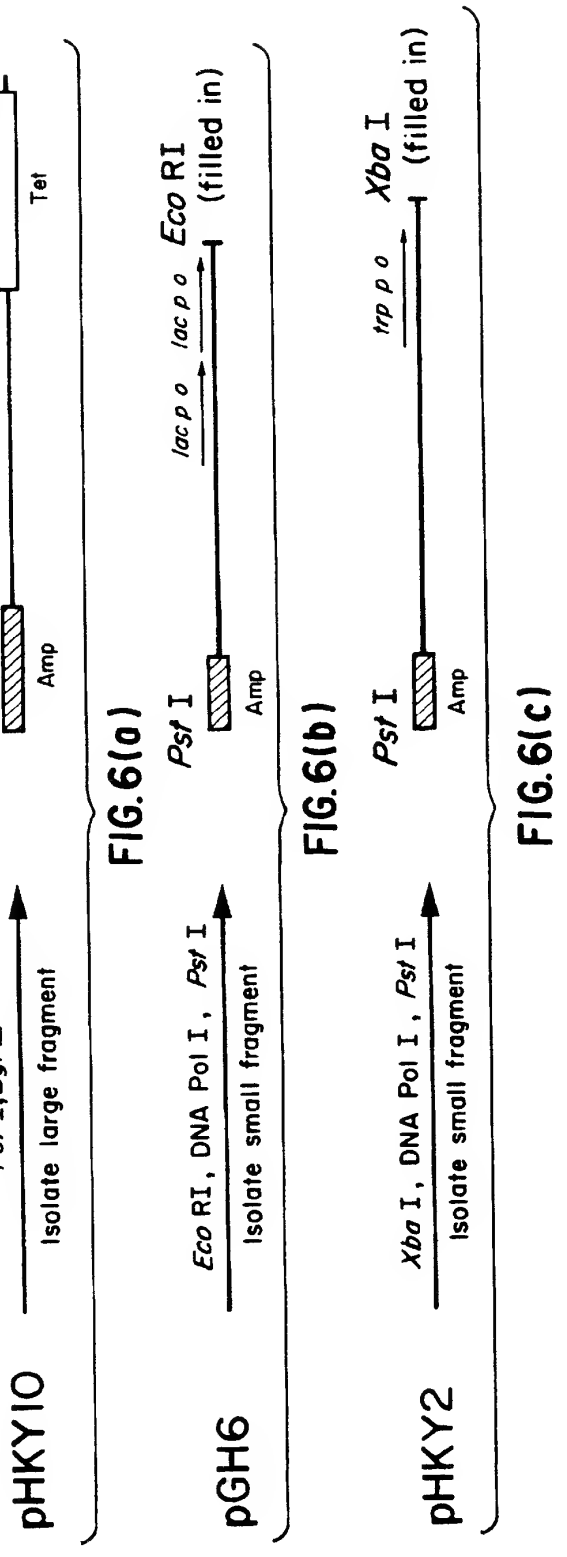

met
C ATG
G TAC
FIF Gene
Bgl II
505 bp
+ (a), (b)
T4 DNA ligase
Eco RI
Pst I
FIF
Bgl II
Tc^R
lac po
lac po
pFIF lac 9
Pst I
Ap^R
met
ATG
TAC
FIF Gene
Bgl II
504 bp
+ (a), (c)
T4 DNA ligase
Xba I
Pst I
FIF
Bgl II
Tc^R
trp po
pFIF trp 69
Pst I
Ap^R

FIG. 6


pHKY10
Pst I, Bgl II
Isolate large fragment
Pst I
Amp
Bgl II
Tet
FIG. 6(a)
pGH6
Eco RI, DNA Pol I, Pst I
Isolate small fragment
Pst I
Amp
lac p o
lac p o
Eco RI
(filled in)
FIG. 6(b)
pHKY2
Xba I, DNA Pol I, Pst I
Isolate small fragment
Pst I
Amp
trp p o
Xba I
(filled in)
FIG. 6(c)

MATURE HUMAN FIBROBLAST INTERFERON

CROSS-REFERENCE TO RELATED APPLICATION

This is a continuation of application Ser. No. 889,722, filed Jul. 28, 1986, now abandoned which is a divisional of application Ser. No. 291,892, now abandoned filed Aug. 11, 1981, which is a continuation-in-part of application Ser. No. 190,799, filed Sep. 25, 1980, now abandoned.

FIELD OF THE INVENTION

This invention relates to the microbial production, via recombinant DNA technology, of human fibroblast interferon for use in the treatment of viral and neoplastic diseases, and to the means and end products of such production.

BACKGROUND OF THE INVENTION

The publications and other materials referred to herein to illuminate the background of the invention and, in particular cases, to provide additional detail respecting its practice are incorporated herein by reference and, for convenience, are numerically referenced in the following text and respectively grouped in the appended bibliography.

Recombinant DNA Technology

With the advent of recombinant DNA technology, the controlled microbial production of an enormous variety of useful polypeptides has become possible. Already in hand are bacteria modified by this technology to permit the production of such polypeptide products such as somatostatin, the (component) A and B chains of human insulin, human growth hormone. More recently, recombinant DNA techniques have been used to occasion the bacterial production of proinsulin, thymosin alpha 1, (an immune potentiating substance produced by the thymus) and leukocyte interferon.

The workhorse of recombinant DNA technology is the plasmid, a non-chromosomal loop of double-stranded DNA found in bacteria and other microbes, oftentimes in multiple copies per cell. Included in the information encoded in the plasmid DNA is that required to reproduce the plasmid in daughter cells (i.e., a "replicon") and ordinarily, one or more selection characteristics such as, in the case of bacteria, resistance to antibiotics which permit clones of the host cell containing the plasmid of interest to be recognized and preferentially grown in selective media. The utility of plasmids lies in the fact that they can be specifically cleaved by one or another restriction endonuclease or "restriction enzyme", each of which recognizes a different site on the plasmidic DNA. Thereafter heterologous genes or gene fragments may be inserted into the plasmid by endwise joining at the cleavage site or at reconstructed ends adjacent to the cleavage site. DNA recombination is performed outside the cell, but the resulting "recombinant" plasmid can be introduced into it by a process known as transformation and large quantities of the heterologous gene-containing recombinant plasmid obtained by growing the transformant. Moreover, where the gene is properly inserted with reference to portions of the plasmid which govern the transcription and translation of the encoded DNA message, the resulting expression vehicle can be used to actually produce the polypeptide sequence for which the inserted gene codes, a process referred to as expression.

Expression is initiated in a region known as the promoter which is recognized by and bound by RNA polymerase. In some cases, as in the tryptophan or "trp" promoter preferred in the practice of the present invention, promoter regions are overlapped by "operator" regions to form a combined promoter-operator. Operators are DNA sequences which are recognized by so-called repressor proteins which serve to regulate the frequency of transcription initiation at a particular promoter. The polymerase travels along the DNA, transcribing the information contained in the coding strand from its 5' to 3' end into messenger RNA which is in turn translated into a polypeptide having the amino acid sequence for which the DNA codes. Each amino acid is encoded by a nucleotide triplet or "codon" within what may for present purposes be referred to as the "structural gene", i.e. that part which encodes the amino acid sequence of the expressed product. After binding to the promoter, the RNA polymerase first transcribes nucleotides encoding a ribosome binding site, then a translation initiation or "start" signal (ordinarily ATG, which in the resulting messenger RNA becomes AUG), then the nucleotide codons within the structural gene itself. So-called stop codons are transcribed at the end of the structural gene whereafter the polymerase may form an additional sequence of messenger RNA which, because of the presence of the stop signal, will remain untranslated by the ribosomes. Ribosomes bind to the binding site provided on the messenger RNA, in bacteria ordinarily as the mRNA is being formed, and themselves produce the encoded polypeptide, beginning at the translation start signal and ending at the previously mentioned stop signal. The desired product is produced if the sequences encoding the ribosome binding site are positioned properly with respect to the AUG initiator codon and if all remaining codons follow the initiator codon in phase. The resulting product may be obtained by lysing the host cell and recovering the product by appropriate purification from other bacterial protein.

Fibroblast Interferon

Human fibroblast interferon (FIF) is an antiviral protein which also exhibits a wide range of other biological activities (see ref. 1 for review). It has reportedly been purified to homogeneity as a single polypeptide of 19,000–20,000 molecular weight having a specific activity of 2 to 10×10^8 units/mg (2,3). The sequence of the 13 NH_2-terminal amino acids of FIF has been determined (4). Houghton et al. (5) have used synthetic deoxyoligonucleotides (predicted from this amino acid sequence) to determine the sequence of the 276 5'-terminal nucleotides of FIF mRNA. Taniguchi et al. (6) and Derynck et al. (7) have recently employed RNA selection procedures to identify cloned cDNA copies of FIF mRNA in *E. coli*. See also Taniguchi et al., *Gene* 10, 11 (1980) and *Proc. Natl. Acad. Sci.* (*U.S.A.*) 77, 5230 (1980) and *Nature* 285, 547 (1980).

While isolation from donor fibroblasts has provided sufficient material for partial characterization and limited clinical studies with homogeneous fibroblast interferon, it is a totally inadequate source for the amounts of interferon heeded for large scale clinical trials and for broad scale prophylactic and/or therapeutic use thereafter. Indeed, presently clinical investigations employing human fibroblast-derived interferon in antitumor and antiviral testing have principally been confined to crude (<1 percent pure) preparations of the material, and long lead times for the manufacture of sufficient quantities, even at unrealistic price levels, have critically delayed investigation on an expanded front.

We perceived that application of recombinant DNA technology would be the most effective way of providing large

quantities of fibroblast interferon which, despite the absence in material so produced of the glycosylation characteristic of human-derived material, could be employed clinically in the treatment of a wide range of viral and neoplastic diseases.

More particularly, we proposed and have since succeeded in producing mature human fibroblast interferon microbially, by constructing a gene therefor which could then be inserted in microbial expression vehicles and expressed under the control of microbial gene regulatory controls.

Our approach to obtaining a fibroblast gene involved the following tasks:

1. Partial amino acid sequences would be obtained by characterization of fibroblast interferon purified to essential homogeneity, and sets of synthetic DNA probes constructed whose codons would, in the aggregate, represent all the possible combinations capable of encoding the partial amino acid sequences.
2. Bacterial colony banks would be prepared containing cDNA from induced messenger RNA. The probes of part (1) would be used to prime the synthesis of radio-labelled single stranded cDNA for use as hybridization probes. The synthetic probes would hybridize with induced mRNA as template and be extended by reverse transcription to form induced, radio-labelled cDNA. Clones from the colony bank that hybridized to radio-labelled cDNA obtained in this manner would be investigated further to confirm the presence of a full-length interferon encoding gene. Any partial length putative gene fragment obtained would itself be used as a probe for the full-length gene.
3. The full-length gene obtained above would be tailored, using synthetic DNA, to eliminate any leader sequence that might prevent microbial expression of the mature polypeptide and to permit appropriate positioning in an expression vehicle relative to start signals and the ribosome binding site of a microbial promoter. Expressed interferon would be purified to a point permitting confirmation of its character and determination of its activity notwithstanding the absence of glycosylation.

SUMMARY OF INVENTION

A series of replicable plasmidic expression vehicles have been constructed which direct the high level synthesis in transformant microorganisms of a mature polypeptide with the properties of authentic human fibroblast interferon. The product polypeptide exhibits the amino acid sequence of such interferon and is active in in vitro testing despite the lack of glycosylation characteristic of the human-derived material. Reference herein to the expression of "mature fibroblast interferon," connotes the bacterial or other microbial production of an interferon molecule unaccompanied by associated glycosylation and the presequence that immediately attends mRNA translation of the human fibroblast interferon genome. Mature fibroblast interferon, according to the present invention, is immediately expressed from a translation start signal (ATG) which also encodes the first amino acid codon of the natural product. The presence or absence of the methionine first amino acid in the microbially expressed product is governed by a kinetic phenomenon dependent on fermentation growth conditions and/or levels of expression in the transformant host. Mature fibroblast interferon could be expressed together with a conjugated protein other than the conventional leader, the conjugate being specifically cleavable in an intra- or extracellular environment. See British Patent Publication No. 2007676A. Finally, the mature interferon could be produced in conjunction with a microbial "signal" peptide which transports the conjugate to the cell wall, where the signal is processed away and the mature polypeptide secreted.

DESCRIPTION OF THE FIGURES

FIG. 1 depicts the protein sequence information used to design degenerate dodecanucleotide primers. Below the protein sequence are corresponding mRNA sequences and six pools of complementary deoxyoligonucleotide primers.

FIGS. **2**A–C are autoradiographs of nitrocellulose-bound plasmid DNA prepared from some of the 600 bacterial transformants having DNA from the fibroblast cDNA library. The plasmid DNA of the nitrocellulose membrane was hybridized with either probe A, B, or C.

FIG. **3** presents DNA sequence of the cDNA insert of clone pFIF3 as determined by the Maxam-Gilbert procedure. The deduced protein sequence in printed above the DNA sequence.

FIG. 4 is a diagram showing the approach used to remove the signal peptide coding regions from pFIF3.

FIG. **5** is an autoradiograph of the polyacrylamide gel used to isolate the 141bp fragment lacking the signal peptide coding regions.

FIG. **6**, and **6***a*, **6***b*, **6***c* schematically depicts the construction of plasmids coding for the direct expression of mature fibroblast interferon. Restriction sites and residues are as shown ("Pst I", etc.). "Ap^R" and "Tc^R" connote portions of the plasmid (s) which express, respectively, ampicillin and tetracycline resistance. The legend "p o" is an abbreviation for "promoter operator."

DESCRIPTION OF THE PREFERRED EMBODIMENTS

A. Microorganisms Employed

The work described involved used of the microorganism: *E. coli* K-12 strain 294 (end A, thi^-, hsr^-, hsm_k^+), as described in British Patent Publication No. 2055382 A. This strain has been deposited on Oct. 28, 1978 with the American Type culture Collection, which is located at 12301 Parklawn Drive. Rockville, Md. 20852, and given ATCC Accession No. 31446. All recombinant DNA work was performed in compliance with applicable guidelines of the National Institutes of Health.

The invention, in its most preferred embodiments, is described with reference to *E. coli*, including not only strain *E. coli* K-12 strain 294, defined above, but also other known *E. coli* strains such as *E. coli* B, *E. coli x* 1776 and *E. coli* W 3110, or other microbial strains many of which are deposited and (potentially) available from recognized microorganism depository institutions, such as the American Type Culture Collection (ATCC)—cf. the ATCC catalogue listing. See also German Offenlegungschrift 2644432. These other microorganisms include, for example, Bacilli such as Bacillus subtilis and other enterobacteriaceae among which can be mentioned as examples Salmonella typhimurium and Serratia marcesans, utilizing plasmids that can replicate and express heterologous gene sequences therein. Yeast, such as Saccharomyces cerevisiae, may also be employed to advantage as host organism in the preparation of the interferon proteins hereof by expression of genes coding therefor under the control of a yeast promoter.

MATERIALS AND METHODS

General methods.

Restriction enzymes were purchased from New England Biolabs and used as directed. Plasmid DNA was prepared by a standard cleared lysate procedure (8) and purified by column chromatography on Biogel A-50M (Bio-Rad). DNA sequencing was performed using the method of Maxam and Gilbert (9). DNA restriction fragments were isolated from polyacrylamide gels by electroelution. DNA fragments were radiolabeled for use as hybridization probes by the random calf thymus DNA priming procedure of Taylor et al. (10). In situ colony hybridizations were performed by the Grunstein-Hogness procedure (11).

Chemical synthesis of deoxyoligonucleotides.

The deoxyoligonucleotides were synthesized by the modified phosphotriester method in solution (12), using trideoxynucleotides as building block (13). The material and general procedures were similar to those described (14). The six pools of primers (Fib 1–6) containing four dodecanucleotides each were obtained by separately coupling two hexamer pools (of two different 5'-terminal sequences each) with three different hexamer pools (of two different 3'-terminal sequences each).

Induction of fibroblasts.

Human fibroblasts (cell line GM-2504A) were grown as described previously (15). Growth medium (Eagles's minimal essential medium containing 10 percent fetal calf serum) was removed from roller bottles (Corning, 850 cm^2) and replaced with 50 ml growth medium containing 50 μg/ml of poly (I):poly (C) (PL Biochemicals) and 10 μg/ml cycloheximide. This induction medium was removed after 4 hours at 37° C. and cell monolayers were washed with PBS (0.14M NaCl, 3 mM KCl, 1.5 mM KH_2PO_4, 8 mM Na_2HPO_4). Each bottle was incubated at 37° C. with 10 ml of a trypsin-EDTA solution (Gibco 610-5305) until cells were detached, and fetal calf serum was added to a concentration of 10 percent. Cells were spun for 15 minutes at 500× g and pellets were resuspended in PBS, pooled, and resedimented. Cells were frozen in liquid nitrogen. Approximately 0.17 g of cells were obtained per roller bottle.

Preparation and assay of interferon mRNA.

Poly extraction and oligo(dT)-cellulose chromatography as described elsewhere (16). The poly (A) containing RNA was enriched for interferon mRNA by centrifugation on a linear 5 percent to 20 percent (w/v) sucrose gradient. The RNA samples were heated to 80° C. for 2 minutes, rapidly cooled, layered over the gradient, and centrifuged for 20 hours at 30,000 rpm at 4° C. in a Beckman SW-40 rotor. Fractions were collected, ethanol precipitated, and dissolved in H_2O.

One microgram samples of mRNA were injected into *Xenopus laevis* oocytes as described previously (17,18). The injected oocytes were incubated 24 hours at 21° C., homogenized, and centrifuged for 5 minutes at 10,000× g. The interferon in the supernatant was determined by the cytopathic effect (CPE) inhibition assay (1) using Sindbis virus and human diploid (WISH) cells. Interferon titers of 1,000 to 6,000 units recovered (NIH reference standard) per microgram of RNA injected were routinely obtained for the 12S species of mRNA.

Synthesis and cloning of cDNA.

Single stranded cDNA was prepared in 100 μl reactions containing 5 μg of 12S fraction mRNA, 20 mM Tris-HCl (pH 8.3), 20 mM KCl, 8 mM $MgCl_2$, 30 mM β-mercaptoethanol, 100 μCi of (α^{32}p)dCTP (Amersham) and 1 mM dATP, dCTP, dGTP, dTTP. The primer was the synthetic Hind III decamer dCCAAGCTTGG (19), which had been extended at the 3' terminus with about 20 to 30 deoxythymidine residues using terminal deoxynucleotidyl transferase (20). 100 units of AMV reverse transcriptase were added and the reaction mixture was incubated at 42° C. for 30 minutes. The second strand DNA synthesis was carried out as described previously (21). The double stranded cDNA was treatcd with 1200 units of S1 nuclease (Miles Laboratories) for 2 hours at 37° C. in 25 mM sodium acetate (pH 4.5), 1 mM $ZnCl_2$, 0.3M NaCl. After phenol extraction the mixture was separated electrophoretically on an 8 percent polyacrylamide gel. cDNA (~0.5 μg) ranging from 500 to 1500 base pairs in size was recovered by electroelution. A 20 ng aliquot was extended with deoxyC residues using terminal deoxynucleotidyl transferase (20), and annealed with 100 ng of pBR322 which had been cleaved with Pst I and tailed with deoxyG residues (20). The annealed mixture was used to transform *E. coli* k-12 strain 294 (22) by a published procedure (23). Strain 294 was used throughout in the work described here, and has been deposited with the American Type Culture collection, accession no. 31446.

Preparation of induced and uninduced ^{32}p-cDNA probes. 5 μg of 12S mRNA were combined with either 2 μg of oligo $(dT)_{12-18}$ (Collaborative Research) or 5 μg of each synthetic primer pool (Fib 1 to Fib 6) in 60 μl of 10 mM Tris-HCl (ph 8), 1 mM EDTA. The mixtures were boiled 3 minutes, and quenched on ice. 60 μl of 40 mM Tris-HCl (pH 8.3), 40 mM KCl, 16 mM $MgCl_2$,60 mM β-mercaptoethanol, 1 mM dATP, dGTP, dTTP and 5×10^{-7}M (α-^{32}p) dCTP (Amersham, 2,000–3,000 Ci/mmole) was added to each template-primer mix at 0° C. After the addition of 100 units of AMV reverse transcriptase, the reactions were incubated at 42° C. for 30 minutes and purified by passage over 10 ml Sephadex G-50 columns. The products were treated with 0.3N NaOH for 30 minutes at 70° C., neutralized, and ethanol precipitated.

The ^{32}p-cDNAs were combined with 100 μg of poly (A) mRNA from uninduced fibroblasts in 50 μl of 0.4M sodium phosphate (pH6.8), 0.1 percent SDS. The mixtures were heated at 98° C. for 5 minutes and allowed to anneal 15 hours at 45° C. The DNA-RNA hybrids. (containing uninduced cDNA sequences) were separated from single-stranded DNA (induced cDNA sequences) by chromatography on hydroxyapatite as described by Galau et al. (24). The DNA-RNA hybrids were treated with alkali to remove RNA.

Screening of recombinant plasmids with ^{32}p-cDNA probes.

Approximately 1 μg samples of plasmid DNA were prepared from individual transformants by a published procedure (25). The DNA samples were linearized by digestion with Eco RI, denatured in alkali, and applied to each of three nitrocellulose filters (Schleicher and Schuell, BA85) by the dot hybridization procedure (26). The filters were hybridized with the ^{32}p-cDNA probes for 16 hours at 42° C. in 50 percent formamide, 10× Denhardt's solution (27), 6×SSC, 40 mM Tris-HCl (pH 7.5), 2 mM EDTA, 40 μg/ml yeast RNA. Filters were washed with 0.1×SSC, 0.1 percent SDS twice for 30' at 42° C., dried, and exposed to Kodak XR-2 x-ray film using Dupont Lightning-Plus intensifying screens at −80° C.

Construction of plasmids for direct expression of FIF.

The synthetic primers I (dATGAGCTACAAC) and II (dCATGAGCTACAAC) were phosphorylated using T4 polynucleotide kinase and (γ-^{32}P)ATP (Amersham) to a specific activity of 700 Ci/mmole as described previously (28). Primer repair reactions were performed as follows: 250 pmoles of the ^{32}P-primers were combined with 8 μg (10 pmole) of a 1200 bp Hha I restriction fragment containing the FIF cDNA sequence. The mixture was ethanol precipitated, resuspended in 50 μl H_2O, boiled 3 minutes, quenched in a dry ice-ethanol bath, and combined with a 50 μl solution

of 20 mM Tris-HCl (pH 7.5), 14 mM $MgCl_2$, 120 mM NaCl, 0.5 mM dATP, dCTP, dGTP, dTTP at 0° C. 10 units of DNA polymerase I Klenow fragment (Boehringer-Mannheim) were added and the mixture was incubated at 37° C. for 4½ hours. Following extraction with phenol/$CHCl_3$ and restriction with Pst I, the desired product was purified on a 6 percent polyacrylamide gel. Subsequent ligations were done at room temperature (cohesive termini) or 4° C. (blunt ends) using previously detailed conditions (21,28).

Assay for interferon expression in *E. coli.*

Bacterial extracts were prepared for IF assay as follows: One ml cultures were grown overnight in LB (29) containing 5 μg/ml tetracycline, then diluted into 25 ml of M9 medium (29) containing 0.2 percent glucose, 0.5 percent casamino acids and 5 μg/ml tetracycline. 10 ml samples were harvested by centrifugation when A_{550} (Absorbance at 500 nanometers) reached 1.0. The cell pellets were quickly frozen in a dry ice-ethanol bath and cleared lysates were prepared as described by Clewell (8). Interferon activity in the supernatants was determined by comparison with NIH FIF standards using cytopathic effect (CPE) inhibition assays as reviewed previously (1). Two different assays were used: (a) WISH (human amnion) cells were seeded in microtiter dishes. Samples were added 16 to 20 hours later and diluted by serial 2-fold dilution. Sindbis virus was added after at least 3 hours in incubation. Plates were stained 20 to 24 hours later with crystal violet. (b) MDBK (bovine kidney) cell line was seeded simultaneously with 2-fold dilutions of samples. Vesicular stomatitis virus was added after 2 to 3 hours incubation and plates were stained with crystal violet 16 to 18 hours later. To test pH 2 stability bacterial extracts and standards were diluted in minimal essential medium to a concentration of 1000 units/ml. One ml aliquots were adjusted to pH 2 with 1N HCl, incubated at 4° C. for 16 hours, and neutralized by addition of NaOH. IF aetivity was determined by the CPE inhibition assay using human amnion cells. To establish antigenic identity 25 μl aliquots of the 1000 U/ml interferon samples (untreated) were incubated with 25 μl of rabbit antihuman leukocyte interferon for 60' at 37° C., centrifuged at 12,000× g for 5 minutes and the supernatant assayed. Fibroblast and leukocyte interferon standards were obtained from the National Institutes of Health. Rabbit antihuman leukocyte interferon was obtained from the National Institute of Allergy and Infectious Diseases.

RESULTS

Chemical synthesis of primer pools complementary of FIF mRNA.

The amino-terminal protein sequenee of human fibroblast interferon (4) permitted us to deduce the 24 possible mRNA sequences whieh could code for the first four amino acids. The 24 complementary deoxyoligonucleotides were synthesized in 6 pools of 4 dodecamers each (FIG. **1**).

The six pools of 4 deoxyoligonucleotides each were synthesized by a modified phosphotriester method that has been used previously for the rapid synthesis of oligonucleotides in solution (12) and on solid phase (14). The basie strategy involved reacting two different 3'-blocked trimers with an excess of a single 5'-protected trimer to yield a pool of two hexamers, each represented equally. The coupling of two pools, each containing two hexamers, then resulted in a pool of four dodecamers.

Identifieation of FIF cDNA clones.

Using 12S mRNA from induced human fibroblasts (1,000 units IF activity per μg in ooeyte assay), double stranded cDNA was prepared and inserted into pBR322 at the Pst I site by the standard dG:dC tailing method (20). A fibroblast cDNA library consisting of 30,000 ampicillin-sensitive, tetracyeline-resistant transformants of *E. coli* K-12 strain 294 was obtained from 20 ng of cDNA ranging in size from 550 to 1300 base pairs. Plasmid DNA was prepared from 600 of the transformants and applied to 3 sets of nitrocellulose filters as described in Materials and Methods.

The approach followed in the identification of hybrid plasmids eontaining fibroblast interferon cDNA sequenees was similar to that used to identify human leukocyte interferon recombinant plasmids (30). Radiolabeled cDNA hybridization probes were prepared using either the 24 synthetic dodecamers or oligo$(dT)_{12\text{-}18}$ as primers and 12S RNA from induced fibroblasts (5000 units/μg in oocytes) as template. The ^{32}P-cDNAs (specific activity >5×10^8 cpm/μg) obtained were hybridized to a large excess of mRNA isolated from uninduced human fibroblasts, and the mRNA-cDNA hybrids were separated from unreacted cDNA by hydroxyapatitc chromatography (24). The single stranded cDNA fractions should be enriched for sequences which are present in induced fibroblasts but absent in uninduced cells, and the mRNA-cDNA hybrids should represent sequences common to both induced and uninduced cells. Approximately 4×10^6 cpm of single stranded cDNA (hybridization probe A) and 8×10^6 cpm of cDNA-mRNA hybrids were obtained using oligo$(dT)_{12\text{-}18}$ primed cDNA; 1.5×10^6 cpm of single stranded (hybridization probe B) and 1.5×10^5 cpm of hybrids were obtained from cDNA primed using synthetic dodecamer pools Fib 1–6. The cDNA-mRNA hybrids from both fractionations were combined, the RNA hydrolyzed by treatment with alkali, and the ^{32}P-cDNA used as hybridization probe C. Many of the 600 plasmid samples hybridized with both probes A and C, indicating that the hybridization reactions between uninduced mRNA and ^{32}P-cDNA (prior to the hydroxyapatite fractionation step) had not gone to completion. However, only one of the 600 plasmids (pF526) hybridized strongly with the specifically primed, induced cDNA probe B (FIG. **2**). Plasmid pF526 also hybridized with the total oligo$(dT)_{12\text{-}18}$ primed, induced cDNA probe A, and failed to give deteetable hybridization to the combined uninduced probe C.

Pst I digestion of pF526 showed the cloned cDNA insert to be about 550 base pairs long, probably too short to contain the entire eoding region for a protein the size of fibroblast interferon. Therefore, a ^{32}P-labeled DNA probe was prepared from this Pst I fragment by random priming with calf thymus DNA (10). This probe was used to screen 2000 individual eolonies from a newly eonstructed fibroblast cDNA library (the new cDNA library was prepared using 12S mRNA from induced fibroblasts having a titer of 6,000 units/ml in the oocyte assay system). Sixteen clones hybridized to the probe. Plasmids prepared from the majority of these released two fragments when cleaved with Pst 1, indicating that the cDNA eontained an internal Pst 1 site. Clone pFIF3 contained the largest cDNA insert, about 800 base pairs. The DNA sequence of the insert was determined by the Maxam-Gilbert procedure (9) and is shown in FIG. 3. The amino acid sequencc of human fibroblast interferon predicted from the nucleotide sequencc is identical to that reported recently by Taniguehi et al. (31) and by Derynck et al. (7) from DNA sequencing of FIF cDNA clones. A precursor or signal peptide of 21 amino acids is followed by a mature interferon polypeptide of 166 amino acids, a stretch of 196 3'-untranslated nucleotides and a poly(A) tail. The NH_2-terminal 20 amino acids of mature FIF have been directly determined by protein microsequeneing and are the same as those predicted from the DNA sequence. The calculated formula molecular weight of mature human fibro-

blast interferon having the 166 amino acids shown in FIG. **3** is about 20,027.

Direct expression of fibroblast interferon.

To express high levels of mature fibroblast interferon in *E. coli*, initiation of protein synthesis must occur at the ATG codon of the mature polypeptide (amino acid 1) rather than at the ATG of the signal peptide (amino acid S1) (FIG. **3**).

Our approach to removing the signal peptide coding regions from pFIF3 is depicted in FIG. **4**. A 1200 bp DNA fragment which contained the entire cDNA insert was isolated from a polyacrylamide gel after digesting pFIF3 with Hha I. Two separate synthetic deoxyoligonucleotide primers, dATGAGCTACAAC(I) and dCATGAGCTACAAC(II), were prepared. Both primers contain the coding sequence for the first four amino acids of mature fibroblast interferon; primer II has an additional C at the 5'-terminus. Primer repair reactions and subsequent ligations were carried out separately for primers I and II, and gave nearly identical results. Therefore, only reactions using primer I are discussed in detail here. The primers were 5'-radiolabeled using (γ-^{32}P)ATP and T4 polynucleotide kinase, combined with the 1200 bp Hha I DNA fragment and the mixture denatured by boiling. Following hybridization of the primer to the denatured Hha I DNA fragment, *E. coli* DNA polymerase I Klenow fragment (33) was used to catalyze the repair synthesis of the plus (top) strand (FIG. **4**). In addition, the associated 3'≧5' exonuclease activity of the Klenow fragment removed the 3'-protruding end from the minus (bottom) strand, leaving a flush end. Analysis of samples of the reaction mixture by polyacrylamide gel electrophoresis indicated that the repair synthesis did not go to completion, but stopped at several discrete sites. Therefore, the entire reaction mixture was treated with Pst I and the desired 141 bp fragment (180,000 Cerenkov cpm; ~0.3 pmole) was purified by polyacrylamide gel electrophoresis (FIG. **5**). Ligation of this fragment to 1 μg (~4 pmole) of the 363 bp Pst I-Bgl II fragment isolated from pFIF3 (FIG. **4**), followed by Bgl II digestion, yielded 50,000 Cerenkov cpm (~0.1 pmole, ~30 ng) of the 504 bp DNA fragment containing the entire coding sequence for mature fibroblast interferon. The same reactions using primer II gave 83,000 cpm (~0.15 pmole, ~50 ng) of 505 bp product.

The construction of plasmids which direct the synthesis of human fibroblast interferon is outlined in FIG. **6**. Separate expression plasmids were constructed which placed FIF synthesis under the control of the *E. coli* lac or trp promoter-operator systems. Both of these systems have proven useful for the direct expression of eukaryotic genes in *E. coli*: human growth hormone has been efficiently synthesized using the lac system (21) and human leukocyte interferon has been produced at high levels using the trp system (30) and *Nature* 287, 411 (1980).

pBRH trp was digested with EcoRI restriction enzyme and the resulting fragment isolated by PAGE and electroelution. EcoRI-digested plasmid pSom 11 (K. Itakura et al., *Science* 198, 1056 (1977); G.B. patent publication no. 2 007 676 A) was combined with the above fragment. The mixture was ligated with T_4 DNA ligase as previously described and the resulting DNA transformed into *E. coli* K-12 strain 294 as previously described. Transformant bacteria were selected on ampicillin-containing plates. Resulting ampicillin-resistant colonies were screened by colony hybridization (M. Gruenstein et al., Proc Nat'l Acad Sci USA 72, 3951–3965 [1975]) using as a probe the trp promoter-operator-containing the above fragment isolated from pBRHtrp, which had been radioactively labelled with p^{32}. Several colonies shown positive by colony hybridization were selected, plasmid DNA was isolated and the orientation of the inserted fragments determined by restriction analysis employing restriction enzymes BglII and BamHI in double digestion. *E. coli* 294 containing the plasmid designated pSOM7Δ2, which has the trp promoter-operator fragment in the desired orientation was grown in LB medium containing 10 μg/ml ampicillin. The cells were grown to optical density 1 (at 550 nM), collected by centrifugation and resuspended in M9 media in tenfold dilution. Cells were grown for 2–3 hours, again to optical density 1, then lysed and total cellular protein analyzed by SDS (sodium dodecyl sulfate) urea (15 percent) polyacrylamide gel electrophoresis (J. V. Maizel Jr. et al., Meth Viral 5, 180–246 [1971]).

Plasmid pBR322 was Hind III digested and the protruding Hind III ends in turn digested with S1 nuclease. The S1 nuclease digestion involved treatment of 10 μg of Hind III-cleaved pBR322 in 30 μl S1 buffer (0.3M NaCl, 1 mM $ZnCl_2$, 25 mM sodium acetate, pH 4.5) with 300 units S1 nuclease for 30 minutes at 15° C. The reaction was stopped by the addition of 1 μl of 30×S1 nuclease stop solution (0.8M tris base, 50 mM EDTA). The mixture was phenol extracted, chloroform extracted and ethanol precipitated, then EcoRI digested as previously described and the large fragment (1) obtained by PAGE procedure followed by electroelution. The fragment obtained has a first EcoRI sticky end and a second, blunt end whose coding strand begins with the nucleotide thymidine.

Plasmid pSom7 Δ2, as prepared above, was Bgl II digested and the Bgl II sticky ends resulting made double stranded with the Klenow polymerase I procedure using all four deoxynucleotide triphosphates. EcoRI cleavage of the resulting product followed by PAGE and electroelution of the small fragment (2) yielded a linear piece of DNA containing the tryptophan promoter-operator and codons of the LE' "Proximal" sequence upstream from the Bgl II site ("LE'(p)"). The product had an EcoRI end and a blunt end resulting from filling in the Bgl II site. However, the Bgl II site is reconstituted by ligation of the blunt end of the above fragment (2) to the blunt end of the above prepared fragment (1). Thus, the two fragments were ligated in the presence of T_4 DNA ligase to form the recirculated plasmid pHKY 10 which was propagated by transformation into competent *E. coli* strain 294 cells.

Plasmid pGM1 carries the *E. coli* tryptophan operon containing the deletion ΔLE1413 (G. F. Miozzari, et al., (1978) *J. Bacteriology* 133, 1457–1466)) and hence expresses a fusion protein comprising the first 6 amino acids of the trp leader and approximately the last third of the trp E polypeptide (hereinafter referred to in conjunction as LE'), as well as the trp D polypeptide in its entirety, all under the control of the trp promoter-operator system. The plasmid, 20 μg, was digested with the restriction enzyme PvuII which cleaves the plasmid at five sites. The gene fragments were next combined with EcoRI linkers (consisting of a self complementary oligonucleotide of the sequence: pCATGAATTCATG) providing an EcoRI cleavage site for a later cloning into a plasmid containing an EcoRI site. The 20 μg of DNA fragments obtained from pGM1 were treated with 10 units T_4 DNA ligase in the presence of 200 pico moles of the 5'-phosphorylated synthetic oligonucleotide pCATGAATCATG and in 20 μl T_4 DNA ligase buffer (20 mM tris, pH 7.6, 0.5 mM ATP, 10 mM $MgCl_2$, 5 mM dithiothreitol) at 4° C. overnight. The solution was then heated 10 minutes at 70° C. to halt ligation. The linkers were cleaved by EcoRI digestion and the fragments, now with EcoRI ends were separated using 5 percent polyacrylamide gel electrophoresis (hereinafter "PAGE") and the three largest fragments

isolated from the gel by first staining with ethidium bromide, locating the fragments with ultraviolet light, and cutting from the gel the portions of interest. Each gel fragment, with 300 microliters 0.1×TBE, was placed in a dialysis bag and subjected to electrophoresis at 100 v for one hour in 0.1× TBE buffer (TBE buffer contains: 10.8 gm tris base, 5.5 gm boric acid, 0.09 gm Na_2EDTA in 1 liter H_2O). The aqueous solution was collected from the dialysis bag, phenol extracted, chloroform extracted and made 0.2M sodium chloride, and the DNA recovered in water after ethanol precipitation. The trp promoter-operator-containing gene with EcoRI sticky ends was identified in the procedure next described, which entails the insertion of fragments into a tetracycline sensitive plasmid which, upon promoter-operator insertion, becomes tetracycline resistant.

Plasmid pBRH1 (R. I. Rodriguez, et al, Nucleic Acids Research 6, 3267–3287 [1979]) expresses ampicilin resistance and contains the gene for tetracycline resistance but, there being no associated promoter, does not express that resistance. The plasmid is accordingly tetracycline sensitive. By introducing a promoter-operator system in the EcoRI site, the plasmid can be made tetracycline resistant.

pBRH1 was digested with EcoRI and the enzyme removed by phenol extraction followed by chloroform extraction and recovered in water after ethanol precipitation. The resulting DNA molecule was, in separate reaction mixtures, combined with each of the three DNA fragments obtained above and ligated with T_4 DNA ligase as previously described. The DNA present in the reaction mixture was used to transform competent *E. coli* K-12 strain 294, K. Backman et al., Proc Nat'l Acad Sci USA 73, 4174–4198 [1976]) by standard techniques (V. Hershfield et al., Proc Nat'l Acad Sci USA 71, 3455–3459 [1974]) and the bacteria plated on LB plates containing 20 μg/ml ampicillin and 5 μg/ml tetracycline. Several tectracycline-resistant colonies were selected, plasmid DNA isolated and the presence of the desired fragment confirmed by restriction enzyme analysis. The resulting plasmid is designated pBRHtrp.

An EcoRI and BamHI digestion product of the viral genome of hepatitis B was obtained by conventional means and cloned into the EcoRI and BamHI sites of plasmid pGII6 (D. V. Goeddel et al., Nature 281, 544 [1979])) to form the plasmid pHS32. Plasmid pHS32 was cleaved with XbaI, phenol extracted, chloroform extracted and ethanol precipitated. It was then treated with 1 μl *E. coli* polymerase 1, Klenow fragment (Boehringer-Mannheim) in 30 μl polymerase buffer (50 mM potassium phosphate pH 7.4, 7 mM $MgCl_2$, 1 mM β-mercaptoethanol) containing 0.1 mM dTTP and 0.1 mM dCTP for 30 minutes at 0° C. then 2 hr. at 37° C. This treatment causes 2 of the 4 nucleotides complementary to the 5' protruding end of the XbaI cleavage site to be filled in: t,**0170**

Two nucleotides, dC and dT, were incorporated giving an end with two 5' protruding nucleotides. This linear residue of plasmid pHS32 (after phenol and chloroform extraction and recovery in water after ethanol precipitation) was cleaved with EcoRI. The large plasmid fragment was separated from the smaller EcoRI-XbaI fragment by PAGE and isolated after electroelution. This DNA fragment from pHS32 (0.2 μg), was ligated, under conditions similar to those described above, to the EcoRI-Taq I fragment of the tryptophan operon (~0.01 μg), derived from pBRHtrp.

In the process of ligating the fragment from pHS32 to the Eco RI-Taq I fragment, as described above, the Taq I protruding end is ligated to the XbaI remaining protruding end even though it is not completely Watson-Crick base-paired: t,**0180**

A portion of this ligation reaction mixture was transformed into *E. coli* 294 cells, heat treated and plated on LB plates containing ampicillin. Twenty-four colonies were selected, grown in 3 ml LB media, and plasmid isolated. Six of these were found to have XbaI site regenerated via E. coli catalyzed DNA repair and replication: t,**0181**

These plasmids were also found to cleave both with EcoRI and HpaI and to give the expected restriction fragments. One plasmid, designated pTrp 14, was used for expression of heterologous polypeptides, as next discussed.

The plasmid pHGH 107 (D. V. Goeddel et al, *Nature,* 281, 544, 1979) contains a gene for human growth hormone made up of 23 amino acid codons produced from synthetic DNA fragments and 163 amino acid codons obtained from complementary DNA produced via reverse transcription of human growth hormone messenger RNA. This gene, though it lacks the codons of the "pre" sequence of human growth hormone, does contain an ATG translation initiation codon. The gene was isolated from 10 μg pHGH 107 after treatment with EcoRI followed by *E. coli* polymerase I Klenow fragment and dTTP and dATP as described above. Following phenol and chloroform extraction and ethanol precipitation the plasmid was treated with BamHI.

The human growth hormone ("HGH") gene-containing fragment was isolated by PAGE followed by electroelution. The resulting DNA fragment also contains the first 350 nucleotides of the tetracycline resistance structural gene, but lacks the tetracycline promoter-operator system so that, when subsequently cloned into an expression plasmid, plasmids containing the insert can be located by the restoration of tetracycline resistance. Because the EcoRI end of the fragment has been filled in by the Klenow polymerase 1 procedure, the fragment has one blunt and one sticky end, ensuring proper orientation when later inserted into an expression plasmid.

The expression plasmid pTrp14 was next prepared to receive the HGH gene-containing fragment prepared above. Thus, pTrp14 was XbaI digest and the resulting sticky ends filled in with the Klenow polymerase I procedure employing dATP, dTTP, dGTP and dCTP. After phenol and chloroform extraction and ethanol precipitation the resulting DNA was treated with BamHI and the resulting large plasmid fragment isolated by PAGE and electroelution. The pTrp14-derived fragment had one blunt and one sticky end, permitting recombination in proper orientation with the HGH gene containing fragment previously described.

The HGH gene fragment and the pTrp14 ΔXba-BamHI fragment were combined and ligated together under conditions similar to those described above. The filled in XbaI and EcoRI ends ligated together by blunt end ligation to recreate both the XbaI and the EcoRI site: t,**0190**

This construction also recreates the tetracycline resistance gene. Since the plasmid pHGH 107 expresses tetracycline resistance from a promoter lying upstream from the HGH gene (the lac promoter), this construction, designated pHGH 207, permits expression of the gene for tetracycline resistance under the control of the tryptophan promoter-operator. Thus the ligation mixture was transformed into *E. coli* 294 and colonies selected on LB plates containing 5 μg/ml tetracycline.

Plasmid pHGH 207 was EcoRI digested and the trp promoter containing EcoRI fragment recovered by PAGE followed by electroelution. Plasmid pBRH1 was EcoRI digested and the cleaved ends treated with bacterial alkaline phosphatase ("BAP") (1 μg, in 50 mM tris pH 8 and 10 mM $MgCl_2$ for 30 min. at 65° C.) to remove the phosphate groups on the protruding EcoRI ends. Excess bacterial

alkaline phosphatase was removed by phenol extraction, chloroform extraction and ethanol precipitation. The resulting linear DNA, because it lacks phosphates on the protruding ends thereof, will in ligation accept only inserts whose complementary stick ends are phosphorylated but will not itself recircularize, permitting more facile screening for plasmids containing the inserts.

The EcoRI fragment derived from pHGH 207 and the linear DNA obtained from pBRH1 were combined in the presence of T_4 ligase as previously described and ligated. A portion of the resulting mixture was transformed into *E. coli* strain 294 as previously described, plated on LB media containing 5 μg/ml of tetracycline, and 12 tetracycline resistant colonies selected. Plasmid was isolated from each colony and examined for the presence of a DNA insert by restriction endonuclease analysis employing EcoRI and XbaI. One plasmid containing the insert was designated pHKY1.

The plasmid pHKY10, described above, is a derivative of pBR322 which contains a Bgl II site between the tetracycline resistance (Tc^R) promoter and structural gene (32). The large DNA fragment isolated after digesting pHKY10 with Pst I and Bgl II therefore contains part of the ampicillin resistance (Ap^R) gene and all of the Tc^R structural gene, but lacks the Tc^R promoter (FIG. **6**). The plasmid pGH6 (21) was digested with Eco RI, the resulting single stranded ends were filled in with DNA polymerase I, and the plasmid was cleaved with Pst I. The small fragment, containing part of the Ap^R gene, a double lac promoter and lac ribosome binding site, but lacking an ATG initiation triplet was isolated. A similar trp promoter fragment, containing the trp leader ribosome binding site, but lacking an ATG sequence (30), may be isolated from pHKY1 described above; see (32) (see FIG. **6**).

The trp fragment just referred to is an analog of the *E. coli* tryptophan operon from which the so-called trp attenuator has been deleted, See *J. Bact.* 133, 1457 (1978), to controllably heighten expression levels. Expression plasmids containing the modified trp regulon can be grown to predetermined levels in nutrient media containing additive tryptophan in quantities sufficient to repress the promoter-operator system, then be deprived of tryptophan so as to derepress the system and occasion the expression of the intended product.

The expression plasmids may be assembled via three part ligation reactions as shown in FIG. **6**. 15 ng (~0.05 pmole) of the assembled FIF gene (504 or 505 bp), 0.5 μg (~0.2 pmole) of the large Pst I-Bgl II fragment of pHKY10 and 0.2 μg (~0.3 pmole) of the appropriate promoter fragment were ligated and the mixture used to transform *E. coli* 294 (22). Plasmid DNA was prepared from individual transformants and analyzed by restriction mapping. Correct joining of the assembled gene to the promoter fragment should restore the Eco RI (lac) or the Xba I (trp) recognition sequences. The majority of the plasmids gave the expected restriction enzyme digestion patterns. Individual clones (12 containing the trp promoter and 12 containing the lac promoter) were grown and extracts prepared for interferon assay as described in Materials and Methods.

When assayed on human amnion (WISH) cells for antiviral activity by the CPE inhibition assay (1) five of the trp transformants were positive (each approximately equivalent); eleven of the lac transformants gave equivalent IF activities. Therefore, one transformant from each series (pFIFlac9 and pFIFtrp69) was selected for further study (Table 1). DNA sequence analysis demonstrated that the desired attachment of promoter to FIF structural gene had occurred in both cases. t,**0210**

The amounts of fibroblast interferon produced by pFIFlac9 and pFIFtrp69 are shown in Table 1. The trp promoter gave a FIF expression level measurably higher than did the lac promoter. In an attempt to further increase FIF expression levels, pFIFtrp69 was cleaved with Eco RI and two 300 base pair Eco RI fragments containing the trp promoter (30) were inserted. The resulting plasmid, pFIFtrp^{3}69, contains three successive trp promoters which read toward the FIF gene. The amount of FIF synthesized by *E. coli* K-12 strain 294/pFIF trp^{3}69 is 4–5 times that produced by pFIF trp 69 (Table 1). This is apparently due to the derepression of the trp promoter which occurs when trp repressor levels are titrated by the multiple copies of the trp operator.

The FIF produced by *E. coli* K-12 strain 294/pFIFtrp69 behaves like authentic human FIF. As shown in Table 2, its antiviral activity is about 30 times greater on human cells than bovine cells. In addition, the bacterially produced FIF is stable to treatment at pH 2 overnight and is not neutralized by rabbit antihuman leukocyte interferon antibodies (Table 3). t,**0220** t,**0221**

Purification

The purification procedure for bacterial derived fibroblast is as follows:

1. Frozen cells are suspended in twelve times volume per weight with sucrose lysis buffer (100 mM Tris-HCl, 10 percent sucrose, 0.2M NaCl, 50 mM EDTA, 0.2 mM PMSF, pH 7.9) containing lysozyme at 1 mg ml^{-1}. The cell suspension is stirred for 1 hour at 4° C. and centrifuged. Fibroblast interferon activity remains in the supernatant.
2. Polyethyleneimine (5 percent v/v) is added to the sonicated supernatant to a final concentration of 0.5 percent (v/v). The solution is stirred for 1 hour at 4° C. and centrifuged. Interferon activity remains in the supernatant.
3. Solid ammonium sulfate is added to the polyethyleneimine supernatant to a final concentration of 50 percent saturation, stirred for 30 minutes at 4° C. and centrifuged. Interferon activity is in the 50 percent pellet.
4. The 50 percent ammonium sulfate pellet is suspended in one half the volume of the 50 percent ammonium sulfate suspension with Phosphate Buffered Saline (20 mM sodium phosphate 0.15M NaCl, pH 7.4). Polyethylene glycol 6000 (50 percent w/v in PBS) is added to a final concentration of 12½ percent (v/v), stirred at 4° C. for 2 hours and centrifuged. Interferon activity is in the pellet. The pellet is suspended in a minimal volume of sucrose lysis buffer and clarified by centrifugation.

This initial extraction procedure results in a purification of fibroblast interferon from 0.001 percent of the total protein to 0.05 percent of the total protein. This material can be further purified to homogeneity by the following column chromatography steps:

5. Afinity chromatography on Amicon Blue B in sucrose lysis buffer.
6. Anion exchange chromatography on QAE Sephadex in sucrose lysis buffer in the absence of 0.2M NaCl.
7. Size exclusion chromatography on Sephadex G-75 in sucrose lysis buffer.
8. Reverse phase high pressure liquid chromatography.

Parenteral Administration

FIF may be parenterally administered to subjects requiring antitumor or antiviral treatment. Dosage and dose rate may parallel that currently in use in clinical investigations of human derived materials, e.g., about $(1–10)\times10^6$ units daily, and in the case of materials of purity greater than 1 percent-

age, likely up to, e.g., 150×10^6 units daily. Dosages of baeterially obtained FIF could be signifieantly elevated for greater effect owning to the essential absence of human proteins other than FIF, which proteins in fibroblast-derived materials may aet as pyrogens, exhibiting adverse effeets, e.g., malaise, temperature elevation, etc.

As one example of an appropriate dosage form for cssentially homogeneous baeterial FIF in parenteral form, 3 mg. FIF of specific activity of, say, 2×10^8 μ/mg may be dissolved in 25 ml. 5 percentage serum albumin (human) - USP, the solution passed through a bacteriologieal filter and the filtered solution aspetically subdivided into 100 vials, each containing 6×10^6 units pure interferon suitable for parenteral administration. The vials are preferably stored in the eold (–20° C.) prior to use.

The compounds of the present invention can be formulated according to known methods to prepare pharmaceutieally useful compositions, whereby the polypeptide hereof is combined in admixture with a pharmaceutically acceptable carrier vehiele. Suitable vehicles and their formulation are described in Remington's *Pharmaceutical Sciences* by E. W. Martin, which is hereby incorporated by reference. Such compositions will contain an effective amount of the interferon protein hereof together with a suitable amount of vehicle in order to prepare pharmaceutically acceptable compositions suitable for effective administration to the host. One preferred mode of administration is parenteral.

BIBLIOGRAPHY

1. Stewart, W. E. II (1979) The Interferon System, Springer, N.Y.
2. Knight, E. Jr. (1976) Proc. Natl. Acad. Sci. U.S.A. 73, 520–523.
3. Berthold, W., Tan, C., and Tan, Y. H. (1978) J. Biol. Chem. 253, 5206–5212.
4. Knight, E. Jr., Hunkapillar, M. W., Korant, B. D., Hardy, R. W. F. and Hood, L. E. (1980) Science 207, 525–526.
5. Houghton, M., Stewart, A. G., Doel, S. M., Emtage, J. S., Eaton, M. A. W., Smith, J. C., Patel, T. P., Lewis, H. M., Porter, A. G., Birch, J. R., Cartwright, T., and Carey, N. H. (1980) Nucleic Acids Res. 8, 1913–1931.
6. Taniguchi, T., Sakai, M., Fujii-Kuriyama, Y., Muramatsu, M., Kobayashi, S., and Sudo, T. (1979) Proc. Jan Acad. B55, 464–469.
7. Derynck, R., Content, J., DeClercq, E., Volckaert, G., Tavernier, J., Devos, R., and Fiers, W. (1980) Nature 285, 542–547.
8. Clewell, D. B. (1972) J. Bacteriol. 110, 667–676.
9. Maxam, A. M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
10. Taylor, J. M., Illemensee, R. and Summer, S. (1976) Biochim. Biophys. Acta 442, 324–330.
11. Grunstein, M. and Hogness, D. S. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 3961–3965.
12. Crea, R., Kraszewski, A., Hirose, T. and Itakura, K. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 5765–5769.
13. Hirose, T., Crea, R. and Itakura, K. (1978) Tetrahedron Letters 28, 2449–2452.
14. Crea, R. and Horn, T. (1980) Nucleic Acids Res. 8, 2331–2348.
15. Pestka, S., McInnes, J., Havell, E. A. and Vileek, J. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 3898–3901.
16. Green, M., Zehavi-Willner, T., Graves, P. N., McInnes, J. and Pestka, S. (1975) Areh. Biochem. Biophys. 172, 74–89.
17. Gurdon, J. B., Lane, C. D., Woodland, H. R. and Marbaix, G. (1971) Nature 233, 177–182.
18. Cavalieri, R. L., Havell, E. A. Vilcek, J. and Pestka, S. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 3287–3291.
19. Scheller, R. H., Dickerson, R. E., Boyer, H. W., Riggs, A. D. and Itakura, K. (1977) Science 196, 177–180.
20. Chang, A. C. Y., Nunberg, J. H., Kaufman, R. J. Erlich, H. A., Schimke, R. T. and Cohen, S. N. (1978) Nature 275, 617–624.
21. Goeddel, D. V., Heyneker, H. L., Hozumi, T., Arentzen, R., Itakura, K., Yansura, D. G., Ross, M. J., Miozzari, G., Crea, R. and Seeburg, P. H. (1979) Nature 281, 544–548.
22. Backman, K., Ptashne, M. and Gilbert, W., (1976) Proc. Natl. Acad. Sci. U.S.A. 73, 4174–4178.
23. Hershfield, V., Boyer, H. W., Yanofsky, C., Lovett, M. A. and Helinski, D. R. (1974) Proc. Natl. Acad. Sci. U.S.A. 71, 3455–3459.
24. Galau, G. A., Britten, R. J. and Davidson, E. H. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 1020–1023.
25. Birnboim, H. C. and Doly, J. (1979) Nucleic Acid Res. 7, 1513–1523.
26. Kafatos, F. C., Jones, C. W. and Efstratiadis, A. (1979) Nucleic Acids Res. 7, 1541–1552.
27. Denhardt, D. T. (1966) Biochem. Biophys. Res. Comm. 23,641.
28. Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K. and Riggs, A. D. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 106–110.
29. Miller, J. H. (1972) Experiments in Molecular Genetics, pp.431–433, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
30. U.S. patent application of David V. Goeddel and Sidney Pestka U.S. Ser. No. 184909, filed Sep. 8, 1980 for Microbial Production of Mature Human Leukocyte Interferon, Attorney docket 1980/104, assigns Genentech, Inc. and Hoffmann-LaRoche, Inc.
31. Taniguchi, T., Ohno, S., Fujii-Kuriyama, Y. and Muramatsu, M. (1980) Gene 10, 11–15.
32. U.S. patent application Ser. No. 06/133,296 filed Mar. 24, 1980 by Dennis G. Kleid et al., assigned to Genentech, Inc.
33. Klenow, H. and Henningsen, I. (1970) Proc. Natl. Acad. Sei. U.S.A. 65, 168–171.

We claim:

1. A composition comprising water and a nonglycosylated polypeptide having the amino acid sequence of a mature human fibroblast interferon, said nonglycosylated polypeptide having a total of 165 or 166 amino acids and said composition being free of any glycosylated human fibroblast interferon.

2. The composition of claim **1**, said nonglycosylated polypeptide having the amino acid sequenee

X-Ser-Tyr-Asn-Leu-Leu-Gly-Phe-Leu-Gln-Arg-Ser-Ser-Asn-Phe-Gln-Cys-Gln-Lys-Leu-Leu-Trp-Gln-Leu-Asn-Gly-Arg-Leu-Glu-Tyr-Cys-Leu-Lys-Asp-Arg-Met-Asn-Phe-Asp-Ile-Pro-Glu-Glu-Ile-Lys-Gln-Leu-Gln-Gln-Phe-Gln-Lys-Glu-Asp-Ala-Ala-Leu-Thr-Ile-Tyr-Glu-Met-Leu-Gln-Asn-Ile-Phe-Ala-Ile-Phe-Arg-Gln-Asp-Ser-Ser-Ser-Thr-Gly-Trp-Asn-Glu-Thr-Ile-Val-Glu-Asn-Leu-Leu-Ala-Asn-Val-Tyr-His-Gln-Ile-Asn-His-Leu-Lys-Thr-Val-Leu-Glu-Glu-Lys-Leu-Glu-Lys-Glu-Asp-Phe-Thr-Arg-Gly-Lys-Leu-Met-Ser-Ser-Leu-His-Leu-Lys-Arg-Tyr-Tyr-Gly-Arg-Ile-Leu-His-Tyr-Leu-Lys-Ala-Lys-Glu-Tyr-Ser-His-Cys-Ala-Trp-Thr-Ile-Val-Arg-Val-Glu-Ile-Leu-Arg-Asn-Phe-Tyr-Phe-Ile-Asn-Arg-Leu-Thr-Gly-Tyr-Leu-Arg-Asn,

wherein X is H or Met.

3. The composition of claim 2, said nonglycosylated polypeptide having a formula molecular weight of about 20,027.

4. The composition of claim 1, 2 or 3, said composition being free of human proteins.

5. The composition of claim 1, 2 or 3, said composition containing a therapeutically effective amount of said nonglycosylated polypeptide and being suitable for parenteral administration.

6. The composition of claim 4, said composition containing a therapeutically effective amount of said nonglycosylated polypeptide and being suitable for parenteral administration.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,460,811

DATED : October 24, 1995

INVENTOR(S) : David V. Goeddel and Roberto Crea

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

In column 5, line 663, please delete (α^{32}p)dCTP and insert (α^{32}P)dCTP.

In column 7, line 16, please delete 500 and insert 550.

In column 10, line 61, please delete GAATCATG and insert GAATTCATG.

In column 11, line 52, please delete t,0170 and insert

5'	CTAGA——	5'	CTAGA——
3'	T——	3'	TCT——

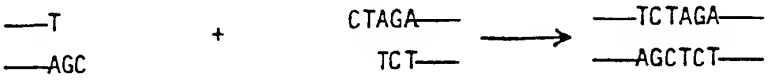
CERTIFICATE OF CORRECTION

PATENT NO. : 5,460,811

DATED : October 24, 1995

INVENTOR(S) : David V. Goeddel and Roberto Crea

In column 11, line 67, please delete **t,0180** and insert



In column 12, line 6, please delete **t,0181** and insert

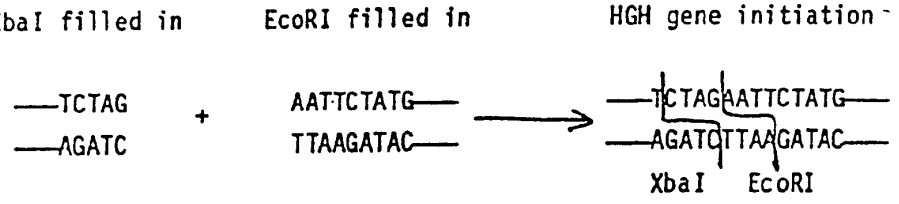
UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,460,811

DATED : October 24, 1995

INVENTOR(S) : David V. Goeddel and Roberto Crea

In column 12, line 51, please delete **t,0190** and insert



UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,460,811

DATED : October 24, 1995

INVENTOR(S) : David V. Goeddel and Roberto Crea

In column 13, line 67, please delete **t,0210** and insert

Table 1. Interferon activity in extracts of E. coli

E. coli K-12 strain 294 transformed by:	Cell density (cells/ml)	IF Activity (units/l culture)	FIF molecules per cell
pBR322	3.5×10^8	-	-
pFIFlac9	3.5×10^8	9.0×10^6	2,250
pFIFtrp69	3.5×10^8	1.8×10^7	4,500
pFIFtrp^{3}69	3.5×10^8	8.1×10^7	20,200

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,460,811

DATED : October 24, 1995

INVENTOR(S) : David V. Goeddel and Roberto Crea

In column 14, line 21, please delete **t,0220** and insert

Table 2. Interferon activities measured on different cell types

	Interferon Activity (units/ml)		
Cells	LeIF	FIF	E. coli K-12 strain 294/pFIF trp69 extract
Human amnion	20,000	10,000	1280
Bovine kidney	13,000	400	40

LeIF and FIF were NIH standard solutions having 20,000 units/ml and 10,000 units/ml respectively. Assays were performed as described in Materials and Methods.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,460,811

DATED : October 24, 1995

INVENTOR(S) : David V. Goeddel and Roberto Crea

In column 14, line 21, please delete t,0221 and insert

Table 3. Comparison of activities of extracts from E. coli K-12 strain 294/pFIFtrp69 with standard human leukocyte and fibroblast interferons

	Interferon Activity (units/ml)		
	LeIF	FIF	E. coli K-12 strain 294/pFIFtrp69
untreated	1000	1000	1000
pH2	1000	1000	1000
rabbit antihuman LeIF antibodies	<16	1000	1000

Experimental procedures described in Materials and methods. Assayed by CPE inhibition using WISH cells/Sindbis virus.

Signed and Sealed this

Ninth Day of April, 1996

Attest:

BRUCE LEHMAN

Attesting Officer

Commissioner of Patents and Trademarks

TAB W

40. Sutcliffe, J. G. *Nucleic Acids Res.* 5, 2721-2728 (1978).
41. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* 74, 560-564 (1977).
42. Seeburg, P. H., Shine, J., Martial, J. A., Baxter, J. O. & Goodman, H. M. *Nature* 270, 486-494 (1977).
43. Proudfoot, N. J. & Brownlee, G. G. *Nature* 263, 211-214 (1976).
44. Fiers, W. *et al. Nature* 273, 113-120 (1978).
45. Soeda, E., Arrand, J. R., Smolar, N., Walsh, J. E. & Griffin, B. E. *Nature* 283, 445-453 (1980).
46. Min Jou, W. *et al. Cell* 19, 683-696 (1980).
47. Shine, J., Seeburg, P. H., Martial, J. A., Baxter, J. D. & Goodman, H. M. *Nature* 270, 494-499 (1977).
48. Blobel, G. *et al. Soc. exp. Biol. Symp.* 33, 9-36 (1979).
49. Habener, J. F. *et al. Proc. natn. Acad. Sci. U.S.A.* 75, 2616-2620 (1978).
50. Oavey, M. W., Sulkowski, E. & Carter, W. A. *J. biol. Chem.* 251, 7620-7625 (1976).
51. Neuberger, A., Gottschalk, A., Marshall, R. D. & Spiro, R. G. in *The Glycoproteins: Their Composition, Structure and Function* (ed. Gottschalk, A.) 450-490 (Elsevier, Amsterdam, 1972).
52. Tomita, M. & Marchesi, V. *Proc. natn. Acad. Sci. U.S.A.* 72, 2964-2968 (1975).
53. Taniguchi, T. *et al. Gene* 10, 11-15 (1980).
54. Taniguchi, T. *et al. Nature*, 285, 547-549 (1980).
55. Mantei, N. *et al. Gene* 10, 1-10 (1980).
56. Roberts, R. J. *Gene* 8, 329-343 (1980).

Human leukocyte and fibroblast interferons are structurally related

Tadatsugu Taniguchi*, Ned Mantei†, Marco Schwarzstein†, Shigekazu Nagata†, Masami Muramatsu* & Charles Weissmann†

*Department of Biochemistry, Cancer Institute, Japanese Foundation for Cancer Research, Tokyo 170, Japan
†Institut für Molekularbiologie I, Universität Zürich, 8093 Zürich, Switzerland

The coding sequences of the cDNAs of cloned human leukocyte interferon I and human fibroblast interferon show homologies of 45% at the nucleotide and 29% at the amino acid level. We conclude that the two genes were derived from a common ancestor.

THE acid-stable human interferons are subdivided into two major groups, fibroblast interferons (F-IF) and leukocyte interferons (Le-IF); these are the major components of the interferons produced by induced fibroblasts and leukocytes, respectively. Some cells, such as the lymphoblastoid Namalva cell line, produce a mixture of 90% Le-IF and 10% F-IF[1,2]. The two interferon types have several features in common: both are glycoproteins with molecular weights ranging from 16,000 to 26,000 (refs 3–9), the induction and shut-off of their synthesis seem to be under similar control[6], and at least some of the responses elicited in target cells are similar, such as induction of an antiviral state, which is accompanied by increased synthesis of several proteins[10–13]. Nonetheless, the two kinds of interferon differ in many respects. Antibodies directed against Le-IF do not neutralize F-IF and vice versa[14], the target cell specificities of the two interferons differ[15], and the sequences of the 13 amino-terminal amino acids of F-IF and Le-IF (from lymphoblastoid cells) show no homology[16,17]. Although Le-IF and F-IF are encoded by different mRNA species[18], it is not known whether these mRNAs are transcribed from distinct genes or whether they arise from the same gene through a common precursor which is processed or spliced in different modes.

We have recently cloned and sequenced one species each of Le-IF (Le-IF I)[19,20] and F-IF cDNA[21,22]. A second species of Le-IF (Le-IF II) cDNA has recently been identified (M. Streuli, S.N. and C.W., unpublished results).

Comparison of the amino acid sequences of Le-IF and F-IF

In Fig. 1 the nucleotide sequences of Le-IF I and F-IF cDNA were aligned so that the AUGs closest to their 5′ termini coincided. From the amino-terminal sequence published for F-IF[16] and lymphoblastoid Le-IF[17], one can determine that for F-IF and Le-IF, respectively, the 21st and 23rd codons following the initiation triplet represent the first amino acid of the interferon polypeptide. Presumably, the stretch in between encodes a signal peptide. As the respective putative signal peptides of Le-IF and F-IF comprise 23 and 21 amino acids, the IF polypeptides, as aligned in Fig. 1, are shifted by two residues relative to their termini. In this alignment, 48 of 166 positions (29%) of the interferon polypeptides have identical amino acids. By introducing appropriate gaps, better homology could be achieved, particularly in the region of the signal sequence; in the present comparison this has not been done.

To plot the degree of homology between the F-IF and Le-IF as a function of the map distance, the sequence was subdivided into segments of 8 amino acids (or 24 nucleotides), each overlapping by 4 amino acids (or 12 nucleotides) with the neighbouring segments, and the per cent coincidence of amino acids (and nucleotides) for each segment was determined (see ref. 23). As seen in Fig. 2, amino acid sequences show three domains of homology. The first one, with the least degree of homology, corresponds to the putative signal sequence, which is rich in hydrophobic residues and has 4 identical amino acid positions out of 21; the second domain, between amino acids 28 and 80 (counted on the Le-IF sequence), has 21 identical residues out of 51 (41% homology), and the third, between positions 115 and 151 (Le-IF sequence), has 19 out of 35 identical residues (54%). The longest stretches of contiguous conserved amino acids are Gln-Phe-Gln-Lys (positions 47–50 of Le-IF and 49–52 of F-IF) and Cys-Ala-Trp (positions 139–141 and 141–143, respectively); the latter sequence is notable because it comprises Cys and Trp, which are preferentially conserved in related proteins[24]. Table I shows that conservation was highest between the interferon polypeptides (not considering the signal sequences) for Trp, Phe, Arg, Cys and Tyr residues, in agreement with the general experience that the amino acids most likely to be conserved between related proteins are Trp > Cys > Tyr > Arg > Phe, His (ref. 24). Even where amino acids are conserved, the codons show one or more nucleotide changes in half the instances. The codons of three out of seven conserved Leu residues are unrelated, as are two of four codons pertaining to conserved Ser residues. This suggests that there is a strong selective pressure favouring the conservation of several amino acids. It is quite likely that at least some of the conserved amino acids are essential for a function common to Le-IF and F-IF, perhaps the induction of the virus-resistant state in the target cell. These findings may provide guidelines for the tailoring of modified[25], possibly shorter polypeptides possessing certain activities of interferon.

Comparison of the nucleotide sequences of Le-IF and F-IF

The nucleic acid sequences show an average homology of 43% in the domain of the signal sequence and of 45% in the interferon polypeptide sequence. On a random basis, about

0028-0836/80/250547-03$01.00

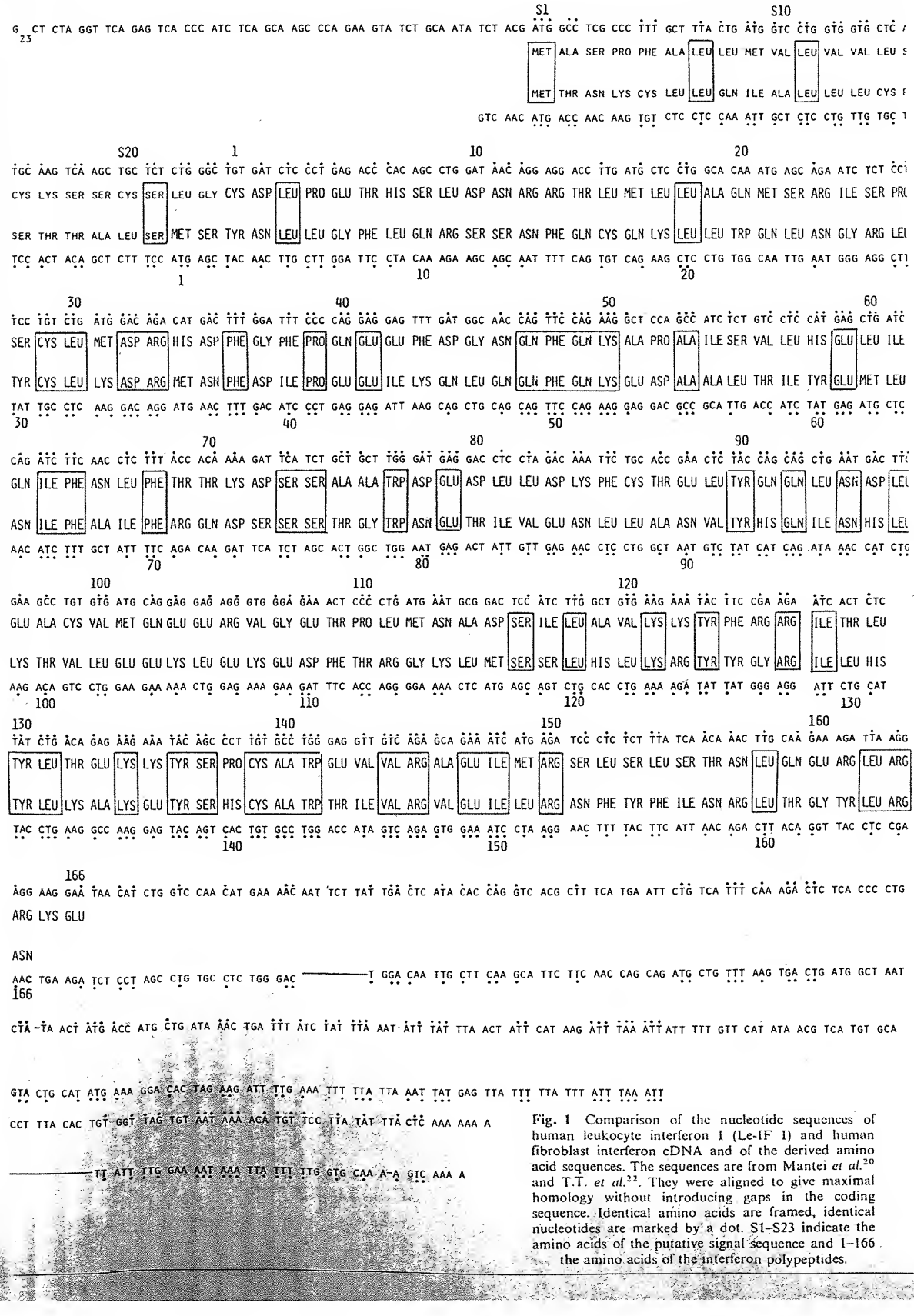


Fig. 1 Comparison of the nucleotide sequences of human leukocyte interferon 1 (Le-IF 1) and human fibroblast interferon cDNA and of the derived amino acid sequences. The sequences are from Mantei *et al.*[20] and T.T. *et al.*[22]. They were aligned to give maximal homology without introducing gaps in the coding sequence. Identical amino acids are framed, identical nucleotides are marked by a dot. S1–S23 indicate the amino acids of the putative signal sequence and 1–166 the amino acids of the interferon polypeptides.

25% of the nucleotide positions should coincide. Within the interferon coding sequence, the nucleotide homologies are more evenly distributed than the amino acid homologies. However, one may distinguish, albeit to a less pronounced degree, the same three blocks of similarity noted for the amino acids. The longest region without mismatches extends for 13 nucleotides (compare 47th to 51st codon of Le-IF with 49th to 53rd codon of F-IF). There are, in addition, sequences of 17, 18 and 20 nucleotides with 3, 3 and 4 mismatches, respectively. The heteropolymeric 3'-terminal noncoding region of Le-IF cDNA has 242 nucleotides, and is longer by 39 residues than its counterpart in F-IF cDNA. In aligning the two sequences, four gaps were introduced to maximize homology, as described by van Ooyen *et al.*[23]. In this way, several segments were matched with 29–41% homology. The introduction of gaps in the alignment may be justified in view of the arguments presented previously, that intervening sequences and noncoding regions of reduplicated genes diverge as a consequence of block insertions and/or deletions in the course of evolution[23,26]. It is unlikely that the extent of homology between Le-IF and F-IF cDNA would allow significant cross-hybridization between the two species.

A common ancestral gene for Le-IF and F-IF

On the basis of our findings, there is no doubt that Le-IF and F-IF genes are derived from a common ancestral sequence. When did the separation of these genes occur? Human α- and β-globin show 57% amino acid mismatches, and human β-globin and myoglobin, as well as α-globin and myoglobin, 76% mismatches. If the rate of divergence of interferons and globins is comparable (however, this is quite uncertain, see ref. 24, p. 50, for proteins showing both higher and lower rates), the separation of interferon genes occurred after that of myoglobin and haemoglobins but before that of α- and β-globins, that is between 500 and 1,000 Myr ago[24], which is about the time vertebrates arose[27]. This would mean that both types of interferon gene should occur in all vertebrates, unless one and/or the other was lost by deletion. Indeed, as shown by the sequencing of 13–24 amino-terminal residues[16,17,28], mouse interferons A and B show significant homology with human fibroblast interferon, and mouse interferon C with human lymphoblastoid interferon, whereas the mouse species A and B on the one hand, and the species C on the other show no significant homology within the short segment sequenced. Thus, at least in the mouse, representatives of both interferon families exist. It will be of interest to determine the evolutionary relationship of these to the third type of interferon, immune or γ-interferon.

After submission of this article, we learnt that Derynck *et al.* had cloned and sequenced fibroblast interferon (see accompanying article[29]), confirming the deduced amino acid sequence of T.T *et al.*[22].

T.T. and M.M. thank Dr H. Sugano for his continuous support and interest. The work at the University of Zürich was supported by Biogen SA and the Schweizerische Nationalfonds.

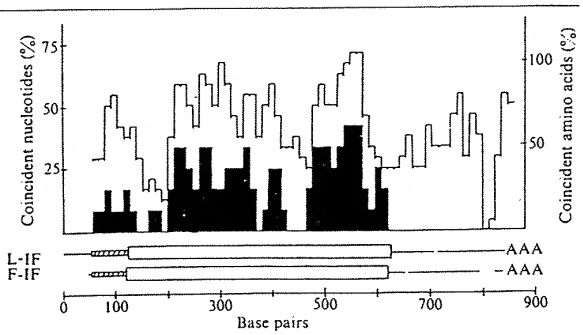


Fig. 2 Similarity of the nucleotide and amino acid sequences of human leukocyte interferon I and fibroblast interferon. The sequences shown in Fig. 1 were subdivided into segments of 8 amino acids or 24 nucleotides, each overlapping by 4 and 12 residues, respectively, with the neighbouring segments. The percentage of concident residues was plotted as a function of map position. Open vertical blocks, nucelotides; filled vertical blocks, amino acids. L-IF, leukocyte interferon cDNA; F-IF, fibroblast interferon cDNA; lines, noncoding sequences; hatched bars, putative signal peptide; open bars, interferon polypeptide.

Table 1 Conservation of amino acids in leukocyte and fibroblast interferon

	F-IF	Le-IF	Conserved amino acids	No. of changes in codon of conserved amino acids 0	1	2	3
Leu	25	22	8	1	4	3	
Cys	3	5	2	1	1		
Asn	12	6	1	1			
Arg	11	12	5	1	3	1	
Phe	9	8	4	2	2		
Pro	1	6	1		1		
Gln	11	10	3	3			
Lys	11	8	3	2	1		
Ala	6	10	2	2			
Glu	13	15	4	4			
Ile	11	7	3	2	1		
Ser	9	13	4		2	1	1
Trp	3	2	2	2			
Tyr	10	4	4	1	3		
Val	5	6	1	1			
Asp	5	11	1	1			
Thr	6	9	0				
Gly	6	3	0				
Met	4	6	0				
His	5	3	0				
Total	166	166	48	24	18	5	1

The data are from T.T. *et al.*[22] and Mantei *et al.*[20].

Received 24 March; accepted 29 April 1980.

1. Havell, E. A., Yip, Y. K. & Vilček, J. *J. gen. Virol.* 38, 51–59 (1977).
2. Paucker, K. *Tex. Rep. Biol. Med.* 35, 23–28 (1977).
3. Bridgen, P. J. *et al. J. biol. Chem.* 252, 6585–6587 (1977).
4. Knight, E. Jr *Proc. natn. Acad. Sci. U.S.A.* 73, 520–523 (1976).
5. Tan, Y. H. *et al. J. biol. Chem.* 254, 8067–8073 (1979).
6. Hayes, T. G., Yip, Y. K. & Vilček, J. *Virology* 98, 351–363 (1979).
7. Rubinstein, M. *et al. Proc. natn. Acad. Sci. U.S.A.* 76, 640–644 (1979).
8. Stewart, W. E. II & Wiranowska-Stewart, M. *Virology* 97, 473–476 (1979).
9. Havell, E. A., Yamazaki, S. & Vilček, J. *J. biol. Chem.* 252, 4425–4427 (1977).
10. Lebleu, B., Sen, G. C., Shaila, S., Cabrer, B. & Lengyel, P. *Proc. natn. Acad. Sci. U.S.A.* 73, 3107–3111 (1976).
11. Hovanessian, A. G., Brown, R. E. & Kerr, I. M. *Nature* 268, 537–540 (1977).
12. Hovanessian, A. G. & Kerr, I. M. *Eur. J. Biochem.* 93, 515–526 (1979).
13. Schmidt, A. *et al. Proc. natn. Acad. Sci. U.S.A.* 76, 4788–4792 (1979).
14. Havell, E. A. *et al. Proc. natn. Acad. Sci. U.S.A.* 72, 2185–2187 (1975).
15. Stewart, W. E. II in *The Interferon System*, 134–145 (Springer, New York, 1979).
16. Knight, E. Jr., Hunkapiller, M. W., Korant, B. O., Hardy, R. W. F. & Hood, L. E. *Science* 207, 525–526 (1980).
17. Zoon, K. C. *et al. Science* 207, 527–528 (1980).
18. Cavalieri, R. L., Havell, E. A., Vilček, J. & Pestka, S. *Proc. natn. Acad. Sci. U.S.A.* 74, 3287–3291 (1977).
19. Nagata, S. *et al. Nature* 284, 316–320 (1980).
20. Mantei, N. *et al. Gene* 10, 1–10 (1980).
21. Taniguchi, T. *et al. Proc. Jap. Acad.* B55, 464–469 (1979).
22. Taniguchi, T., Ohno, S., Fujii-Kuriyama, Y. & Muramatsu, M. *Gene* 10, 11–15 (1980).
23. van Ooyen, A., van den Berg, J., Mantei, N. & Weissmann, C. *Science* 206, 337–344 (1979).
24. Dayhoff, M. O. in *Atlas of Protein Sequence and Structure* Vol. V (Natn. Biomed. Res. Fdn, Washington, 1972).
25. Müller, W., Weber, H., Meyer, F. & Weissmann, C. *J. molec. Biol.* 124, 343–358 (1978).
26. Konkel, D. A., Maizel, J. V. Jr & Leder, P. *Cell* 18, 865–873 (1979).
27. Kurtén, B. in *Proc. XVI int. Congr. Zool.* Vol. 6 (ed. Moore, J. A.) 327–354 (Natural History Press, New York, 1965).
28. Taira, H. *et al. Science* 207, 528–530 (1980).
29. Derynck, R. *et al. Nature* 285, 542–547 (1980).

TAB X

Proc. Natl. Acad. Sci. USA
Vol. 76, No. 11, pp. 5631–5635, November 1979
Biochemistry

Detection of proteins like human γ and β globins in *Escherichia coli* carrying recombinant DNA plasmids

(protein expression/cDNA cloning/radioimmunoassay)

LOIS B. WILSON, HERNAN E. GRENETT, DANIEL M. DORAN, FRED A. GARVER, AND JOHN T. WILSON

Department of Cell and Molecular Biology, Medical College of Georgia, Augusta, Georgia 30912

Communicated by Frank H. Ruddle, August 16, 1979

ABSTRACT *Escherichia coli* strain χ1776 carrying recombinant DNA plasmids containing cDNA copies of human β or γ globin mRNAs has been shown by radioimmunoassay to synthesize polypeptides antigenically related to the β and γ chains of human hemoglobin. The γ and β polypeptides have been enriched from lysates on immunoabsorbent columns containing hemoglobin antibodies and shown to specifically inhibit the antigen–antibody binding between ^{125}I-labeled hemoglobin and the homologous antibody but not other hemoglobin–antihemoglobin reactions. Clone JW151, which is known to contain a complete copy of the coding portion of the γ globin mRNA, has been shown to produce a protein that reacts specifically with antibody to the γ chain of fetal hemoglobin, hemoglobin Kenya, and hemoglobin Bart's.

Although transcription and translation of cloned eukaryotic genes in a prokaryotic system have been well documented (1–7), one aspect of molecular cloning that has eluded most efforts is the expression of cloned higher eukaryotic genes in a prokaryotic system. We reasoned that part of the difficulty was technical, and that a sensitive assay, such as radioimmunoassay, would be necessary to detect what might be small amounts of protein production. Recently two such experiments have been reported. Villa-Komaroff *et al.* (8) have reported using immunological techniques to detect the production of rat proinsulin cloned in *Escherichia coli* χ1776, using a cDNA copy of rat pre-proinsulin mRNA inserted in bacterial plasmid pBR322. Also, Mercereau-Puijalon *et al.* (9) have reported the synthesis of chicken ovalbumin-like protein in *E. coli* K-12 harboring a recombinant DNA plasmid in vector pBR322. More than 2 years ago, we produced cDNA clones of human α, β, and γ, globin mRNA (10, 11) inserted into bacterial plasmid pMB9. Here we report the expression of proteins from human cDNA in *E. coli*, by demonstrating that these proteins contain antigenic determinant sites specific for human γ chain and proteins that react with antibodies to human β chain. This research shows that human gene sequences can be expressed in the bacterial cell and, furthermore, that bacterial plasmid pMB9 can be used as a vector for expression as well as pBR322.

MATERIALS AND METHODS

Bacterial Strains, Plasmids, and Media. *E. coli* strain χ1776, carrying a recombinant DNA plasmid, was grown in enriched medium (Luria broth) according to Curtiss *et al.* (12). Bacterial cultures were grown at 37°C to midlogarithmic phase. Preparation of plasmid DNA has been described (10, 11). All growth experiments were performed in a P3 physical containment facility in compliance with National Institues of Health guidelines for recombinant DNA research (13).

Preparation of Crude Lysates. Bacterial cells were harvested by centrifugation at 1600 × *g* for 10 min at 5°C in an IEC B20A centrifuge, resuspended in phosphate-buffered saline containing Trasylol (Mobay Chemical, New York) at 60 units/ml and phenylmethylsulfonyl fluoride (Sigma) at 30 μg/ml, and sonicated, in a biological safety cabinet located in the P3 facility, with an Artek Sonic 300 dismembrator for 3 min or until 90% of bacterial cells were disrupted as measured by the reduction in optical density at 590 nm and by appearance in the light microscope. Cellular debris was removed by centrifugation at 500 × *g* for 10 min, followed by a second centrifugation of the supernatant at 10,000 × *g* for 45 min. The crude lysate was concentrated by negative pressure dialysis against 0.1 M sodium borate buffer containing 0.075 M NaCl and 0.01 M sodium azide, pH 8.3. After concentration, proteins were fractionated by ammonium sulfate precipitation by the addition of an equal volume of 78% saturated ammonium sulfate. After incubation at room temperature for 30 min the supernatant was collected after centrifugation at 1600 × *g* for 30 min at 20°C, and the precipitate was resuspended in distilled water. Both the precipitate and the supernatant fractions were dialyzed twice at 4°C against 0.01 M ammonium bicarbonate and concentrated by lyophilization. The proteins were redissolved in borate buffer.

Isolation and Purification of Hemoglobins. Hemoglobins A ($\alpha_2\beta_2$) and $A_2(\alpha_2\delta_2)$ were isolated from blood of normal donors, Hb F ($\alpha_2\gamma_2$) and Hb Bart's (γ_4) from cord blood, Hb F Malta I ($\alpha_2\gamma_2^{117\ His\rightarrow Arg}$), Hb G Philadelphia ($\alpha_2^{68\ Asn\rightarrow Lys}\beta_2$), Hb S ($\alpha_2\beta_2^{6\ Glu\rightarrow Val}$), and Hb Kenya from individuals known to carry these variants. Hb Kenya has a fused non-α Hb chain consisting of an NH_2-terminal corresponding to the γ chain (to approximately amino acid 86) and the remaining COOH-terminal portion identical to the β chain. Erythrocyte lysates were prepared and hemoglobins were isolated by DEAE-cellulose chromatography as described (14, 15). The dilute hemoglobin fractions were concentrated by ultrafiltration with an Amicon ultrafiltration cell (PM10 Membrane, Amicon, Lexington, MA). Homogeneity of the preparations was analyzed by starch gel electrophoresis (15).

Preparation and Absorption of Antisera. New Zealand White rabbits were immunized with the purified hemoglobins according to the procedure of Garver *et al.* (16).

Crossreacting antibodies were eliminated by immunoadsorption with hemoglobin coupled to AH-Sepharose as reported (17, 18). The specificity of the absorbed antisera was analyzed by measuring the precipitation of both the homologous ^{125}I-labeled hemoglobin and the ^{125}I-labeled hemoglobin that was used for absorption. The antisera were further characterized by inhibition tests using unlabeled homologous and heterologous hemoglobin antigens.

The preparation and specificity of antisera utilized for radioimmunoassay is listed below and described in detail in the accompanying references: (*a*) anti-Hb F serum absorbed with Hb A (γ chain specific) (16); (*b*) anti-Hb A_2 serum absorbed

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

with Hb A (δ chain specific) (16); (*c*) anti-Hb F Malta I serum absorbed with Hb F (specific for Arg at position 117 of the γ chain) (19); (*d*) anti-Hb G Philadelphia serum absorbed with Hb A (specific for Lys at position 68 of the α chain) (17); (*e*) anti-Hb Kenya serum absorbed with Hb A (specific for the region of the non-α chain of Hb Kenya) (18); (*f*) anti-Hb Bart's absorbed with Hb A (specific for γ chain) (unpublished data); and (*g*) anti-Hb S absorbed with Hb A (specific for Val at position 6 of the β chain) (17).

Preparation of Labeled Antigens and Radioimmunoassay. Hemoglobins were labeled with ^{125}I by the chloramine-T procedure (20). Unbound ^{125}I was eliminated by gel filtration and dialysis (16), and the preparations were stored at 4°C.

A standard radioimmunochemical inhibition test (16, 20) was used to determine the presence of immunoactive proteins from crude lysates of bacterial cultures. In these tests, proteins obtained from bacterial cultures were assayed for their ability to inhibit an ^{125}I-labeled hemoglobin–antihemoglobin precipitation reaction.

Isolation of Globin-Chain-Like Proteins. Antigenically active protein was isolated by affinity chromatography. The ammonium sulfate supernatants of lysates were chromatographed on antibody-coupled Sepharose columns. The antibody had been isolated from the antisera by chromatography on hemoglobin-conjugated Sepharose columns, and the retained antibody had been dissociated with 0.05 M glycine·HCl buffer, pH 3.5. After preparation of the antibody affinity column, the columns were washed extensively with 3 M sodium thiocyanate, borate buffer, then glycine·HCl buffer. After washing, the columns were reequilibrated with borate until the absorbance at 280 nm was less than 0.01 prior to the addition of ammonium sulfate supernatant of the bacterial cell lysate. Bound protein was eluted from the columns by the addition of glycine buffer, collected, concentrated, and stored at 4°C.

Gel Electrophoresis and Restriction Endonuclease Digestion. Procedures for restriction endonuclease digestion and gel electrophoresis have been described (10, 11). All restriction enzymes were purchased from New England Bio-Labs or Bethesda Research Laboratories, Rockville, MD.

RESULTS

Screening of Clones. Previously we reported (11) the production of 54 hybrid DNA clones containing sequences complementary to human, α, β, or γ chain globin mRNA. To facilitate the determination of clones as candidates for the production of immunoactive globin proteins, crude lysates of these clones were prepared from 30-ml cultures. Supernatants obtained from each of these clones were tested by radioimmunoassay for the inhibition of the precipitation of ^{125}I-labeled hemoglobin by antibody. Clones that were previously determined (11) to contain sequences complementary to γ globin mRNA were tested in a radioimmunochemical assay against Hb F, whereas clones containing sequences complementary to β globin mRNA were tested in an assay against Hb A. Crude lysates from the majority of clones tested gave no inhibition, and neither did the control of *E. coli* carrying pMB9 plasmid (Table 1). Cultures showing less than 5% inhibition were scored as negative. Two clones, JW151 and JW109, showed slight inhibition with ^{125}I-labeled Hb F (Table 1, line 1) and ^{125}I-labeled Hb A (Table 1, line 3) respectively. These results indicated that these clones would be likely candidates for further study. Because such a low level of inhibition probably represents a low level of immunoactive protein present, experiments were conducted for enrichment of such proteins.

Protein Enrichment. Proteins were fractionated from cultures of JW151 and JW109 by ammonium sulfate precipitation. By this procedure, 70% of the total bacterial protein was precipitated. The ammonium sulfate precipitate and supernatant were tested by radioimmunoassay for inhibition. Because 70% of the antigenic activity was found in the supernatant, this fraction was used for further pruification.

Additional purification of antigenically actice proteins was obtained by the use of affinity chromatography with antibody coupled to Sepharose. Protein from JW 151 ammonium sulfate supernatant (63 mg) was chromatographed on such a column, as illustrated in Fig. 1. This procedure yielded 7% of the total protein (4.4 mg) being eluted by glycine·HCl buffer, after having been bound by Hb F antibody. This represents 2% of the protein recovered from the crude lysate. Similarly, 1.5 mg of protein was eluted from JW109 lysate by using a Sepharose-bound anti-Hb A column. Eluted proteins were further tested by radioimmunoassay for inhibition. In such tests, 50 μg of isolated protein from JW151 gave 82.8% inhibition (Table 1, line 2) as compared to 97.3% inhibition with 0.5μg of purified Hb F. Bound protein (50 μg) from JW109 yielded 75% inhibition (Table 1, line 4), whereas 5 μg of Hb A gave 91.3% inhibition. Lysates of bacterial cultures carrying pMB9 plasmid showed that no immunologically active proteins were retained on the Sepharose-antibody columns (data not shown).

Fig. 2 represents the inhibition curve obtained by radioimmunoassay for enriched globin specific proteins from JW151 at various concentrations. As can be seen, proteins isolated from this clone showed a significant amount of inhibition. However, maximum inhibition obtained for protein from JW151 was 82% (at inputs of greater than 100 μg of protein), whereas Hb F controls yielded 100% inhibition with as little as 5 μg of Hb F. This implies that the enriched protein from JW151 has reduced antigenic activity (on a weight basis) compared to Hb F controls but still exhibits a substantial degree of γ chain antigenic specificity as demonstrated below.

Table 1. Radioimmunoassay for globin chains

Controls					Clone JW151		Clone JW109	
Antiserum (AS)	Normal serum (NS)	Hb F (5.0 μg)	Hb A (5.0 μg)	*E. coli* carrying pMB9 plasmid, crude lysate (100 μl)	Crude lysate (100 μl)	Isolated protein (50 μg)	Crude lysate (100 μl)	Isolated protein (50 μg)
γ chain								
1. 0% (4185)	100% (548)	90.1% (909)		2.3%	11.1% (3781)			
2. 0% (2776)	100% (580)	97.3% (645)				82.8% (957)		
β chain								
3. 0% (2193)	100% (397)		82.4% (714)	2.3%			11.7% (1983)	
4. 0% (2800)	100% (469)		91.3% (672)					75.0% (1050)

Values represent percent inhibition of the precipitation of ^{125}I-labeled hemoglobin by antibody; the actual cpm is given in parentheses. Antiserum (AS) control values represent maximum precipitation in absence of inhibitor. Normal serum (NS) control values (no antibody) represent nonspecific precipitation and therefore reflect expected values for 100% inhibition. Percent inhibition (%I) was calculated by the formula %I = 100 × [(AS − precipitated)/(AS − NS)]. Crude lysates of hybrid clones and pMB9 contained approximately equal protein concentrations (80 mg/ml).

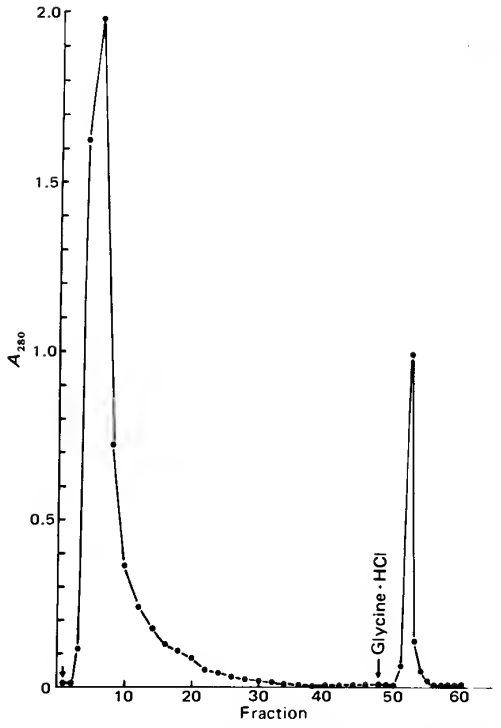


FIG. 1. Protein fractionation on antibody-coupled Sepharose columns. Protein from the ammonium sulfate supernatant of the crude lysate from JW151 was applied to the Sepharose column equilibrated in borate buffer (arrow on left), and the column was washed until the A_{280} was less than 0.01. Bound protein (fractions 50–60) was eluted from the column by the application of glycine-HCl buffer, pH 3.5.

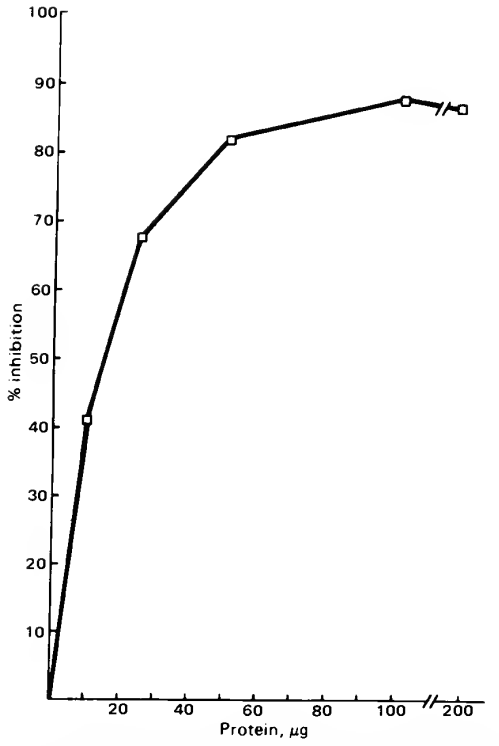


FIG. 2. Inhibition assay for protein obtained from Sepharose column. Protein obtained from JW151 was assayed, at various concentrations, for its ability to inhibit the immunochemical reaction of ^{125}I-labeled γ globin (0.1 μg) and anti-γ globin.

Analysis of Specificity of Globin Protein. In order to determine if enriched protein from JW151 nonspecifically inhibited the antigen–antibody interaction, Sepharose-isolated protein was analyzed in radioimmunoassays with various hemoglobin-specific antibodies. Table 2 illustrates the results of these tests. The γ globin-like protein from JW151 gave no inhibition in radioimmunoassays specific for normal δ chains (Hb A_2), in an assay specific for a γ chain variant (Hb F Malta I), or an α chain variant (Hb G Philadelphia), but gave inhibition in assays for γ chains (Hb F, Hb Bart's). In addition, the purified JW151 protein partially blocked the binding between ^{125}I Hb Kenya and antibody to Hb Kenya, providing further evidence of γ chain antigenic determinants.

The β globin-like protein isolated from JW109 lysates failed to show any inhibition in the radioimmunoassay specific for γ chains. This protein also failed to show any inhibition in the radioimmunoassay for Hb S.

Determination of the Direction of Globin Gene Insertion in pMB9. The orientation of globin complementary sequences inserted within a bacterial plasmid could be of importance for the expression of these sequences within the cell, because their expression will depend on the plasmid and bacterial transcriptional mechanisms. For this reason pMB9 and various recombinant DNA plasmids were digested with restriction endonucleases to determine the orientation of the inserted globin cDNAs within the plasmids. Fig. 3 shows a detailed restriction endonuclease cleavage map of the α, β, and γ globin cDNA as determined from nucleotide base sequence analysis of recombinant DNA plasmids (unpublished data). As can be seen from Fig. 3, β and ^{G}γ globin cDNA contains no *Hha* I enzyme cleavage site. Furthermore, both β globin and ^{G}γ globin cDNAs contain a single *Eco*RI and *Bam* HI enzyme cleavage site. The globin cDNAs were inserted into pMB9 at the *Eco*RI cleavage

Table 2. Inhibition of antigen–antibody reactions

Reactants		% inhibition	
Antigen	Antibody	JW151 bound peak (25 μg)	Control (5 μg)
^{125}I-Hb A_2	Anti-Hb A_2	0	97.0 (Hb A_2)
^{125}I-Hb F	Anti-Hb F (γ chain specific)	67.9	97.3 (Hb F)
^{125}I-Hb F Malta I	Anti-Hb F Malta I ($\gamma^{117\ Arg}$ specific)	0	76.7 (Hb F Malta I)
^{125}I-Hb Kenya	Anti-Hb Kenya (specific for γ portion of non-α chain of Hb Kenya)	44.2	100 (Hb Kenya)
^{125}I-Hb G Philadelphia	Anti-Hb G Philadelphia (specific for $\alpha^{68\ Lys}$)	0	93 (Hb G Philadelphia)
^{125}I-Hb Bart's	Anti-Hb Bart's	32	53 (Hb Bart's)

Test of chain specificity of protein obtained from clone JW151. Values are percent inhibition of antibody precipitation of ^{125}I-labeled hemoglobin (0.1 μg) by Sepharose-isolated protein from JW151. Controls consisted of identical reactions in which unlabeled hemoglobin, specific for the antibody, was added instead of protein from JW151.

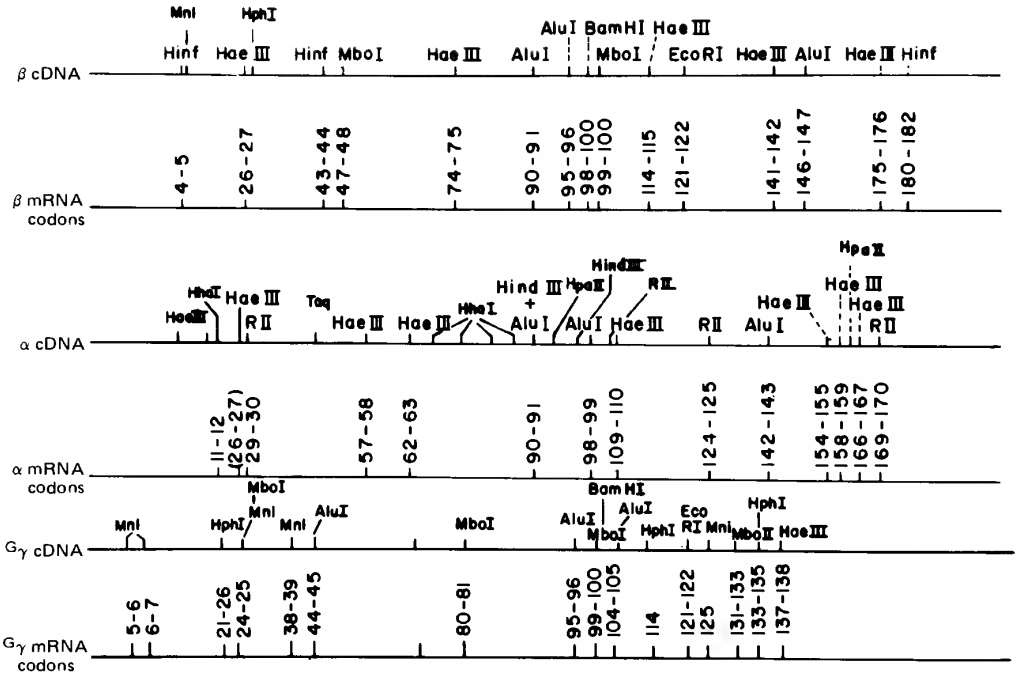


FIG. 3. Restriction endonuclease cleavage maps for the α, β, and $^{G}\gamma$ globin genes, as determined from base sequence analysis of recombinant DNA plasmids (data not shown).

site (10, 11). This site is flanked by two *Hha* I cleavage sites (21), one approximately 900 bases to the left of the *Eco*RI site and the other approximately 400 bases to the right. Previous studies have shown that this *Hha* I restriction fragment of pMB9 (Fig. 4) is the largest fragment obtained by digestion of pMB9 DNA with *Hha* I (11) and contains all of the inserted β and γ globin cDNAs. Therefore, for simplicity, the remaining restriction fragments produced by digestion of pMB9 by *Hha* I will not be considered in this report. Digestion of β and $^{G}\gamma$ globin cDNA plasmids with *Hha* I and *Eco*RI enzymes shows this fragment to be cut into two fragments of unequal size. Further digestion with enzyme *Bam*HI determines the direction of insertion of the globin cDNA. If the largest *Hha* I *Eco*RI fragment is cut by *Bam*HI, then the *Bam*HI site must be associated with the restriction fragment that contains 900 bases complementary to pMB9. Therefore, the cDNA would be inserted 5′ to 3′ with respect to the globin mRNA and with a clockwise orientation from the RI site to the tetracycline site in the plasmid DNA. If the smallest fragment is cut by *Bam*HI, the cDNA would be in a 3′-to-5′ orientation.

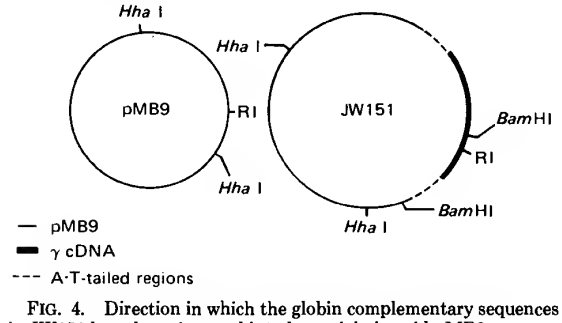


FIG. 4. Direction in which the globin complementary sequences in JW151 have been inserted into bacterial plasmid pMB9.

Fig. 4 shows that plasmid JW151 has been determined to have the γ globin cDNA inserted in a 5′-to-3′ direction. Restriction endonuclease digestion of β globin plasmids JW109 and JW102 has shown that JW109 plasmid has the β globin sequences inserted into pMB9 in the same orientation as the γ globin sequences in JW151 and that JW102 (which has shown no antigenically active protein production) is in the opposite orientation.

DISCUSSION

Garver *et al.* (16) have reported the preparation of monospecific antibodies for the δ and γ chains of human hemoglobin and the incorporation of these chain-specific antisera into radioimmunoassays to quantitate hemoglobins A_2 and F. These workers have also described the methods for preparing antibodies that specifically recognize the single amino acid substitution in 15 variant human hemoglobins (17). Furthermore, Reichlin *et al.* (22) have shown that antibodies produced against native hemoglobin react with free α and β chains, and antibodies against the individual native subunits react with native hemoglobin. Therefore, antibodies produced against native hemoglobin should recognize antigenic sites on free chains synthesized in *E. coli*. By use of such antibodies, we have been able to demonstrate proteins, in our recombinant DNA clones of *E. coli*, showing antigenic activity. The protein production of JW151 has been determined to be γ specific by use of various hemoglobin antigen–antibody systems. Because protein isolated from JW151 showed no interaction with antibody to the δ chain of Hb A_2 but did show reactivity with antibody to the γ chain of both Hb F and Hb Bart's, the protein produced must be γ specific and not α or δ chain specific. Furthermore, because the protein exhibited close to 45% inhibition with Hb Kenya (55% γ, 45% β), then this protein probably contains most of the antigenic determinant sites associated with the NH_2-terminal end of the γ globin chain. As expected, this protein also failed to

show any inhibition with an antibody against Hb F Malta I that specifically recognizes the His → Arg substitution at amino acid residue 117 in the γ chain of Hb F Malta I. Nor did the protein inhibit an antibody against Hb G Philadelphia that is specific for the Asn → Lys substitution at position 68 of the α chain.

Through the use of radioimmunoassays we were able to obtain no more than 85% inhibition with γ-like (Fig. 2) and 75% inhibition with β-like protein preparations (data not shown). This may be indicative of a modified protein that lacks antigenic determinant sites found on normal γ or β chains. This may be attributed to an altered conformation of the protein due to the lack of globin specific sequences or to the transcription of excess plasmid DNA. DNA sequencing studies of the insert in JW151 have shown this hybrid plasmid to contain sequences complementary to the human γ globin mRNA from preceding the initiator codon to sequences in the 3′ untranslated region following the termination codon (unpublished data). With a complete copy of the translated region of the γ mRNA present, one would expect any globin-like protein produced to be complementary to the γ globin chain. Therefore, the decreased antigenicity may be due to the production of a hybrid protein that consists of both bacterial and globin sequences. This can be envisioned if the bacterial promoter for the transcription of this region of pMB9 plasmid initiates transcription upstream from the site of the insert, as is the case for the rat proinsulin system (8). The molecular weight of such proteins will not be that of purified γ or β globin chains, because it will also be dependent on the extent of plasmid sequences that are co-transcribed with the inserted globin sequences.

The ability of cloned eukaryotic genes to be expressed in a prokaryotic system is probably governed by a multitude of factors. One such factor may be the orientation of the inserted gene with respect to the direction of transcription of the plasmid DNA. In our studies, it has been determined that both recombinant plasmids that show expression also have the cDNA inserted into the plasmids in the same orientation. This may be indicative that this orientation is necessary for proper strand specificity during RNA transcription and may therefore further support the idea that the eukaryotic gene is using the prokaryotic mechanisms for transcription.

Note: Since the submission of this paper, Goeddel *et al.* (23) have reported the expression of synthetic human insulin genes cloned in *E. coli* by using plasmid pBR322.

This work was supported in part by National Institutes of Health Grants HLB 15158 and HLB 23294.

1. Struhl, K., Cameron, J. R. & Davis, R. W. (1977) *Proc. Natl. Acad. Sci. USA* **73**, 1471–1475.
2. Struhl, K. & Davis, R. W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5255–5259.
3. Vapnek, D., Hautala, J., Jacobson, J., Giles, N. & Kushner, S. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3058–3062.
4. Rambach, A. & Hogness, D. S. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5041–5045.
5. Ratzlcin, B. & Carbon, J. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 487–491.
6. Meagher, R. B., Tait, R. C., Betlach, M. & Boyer, H. W. (1977) *Cell* **10**, 521–526.
7. Clarke, L. & Carbon, J. J. (1978) *J. Mol. Biol.* **120**, 517–532.
8. Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3727–3731.
9. Mercereau-Puijalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, F. & Kourilsky, P. (1978) *Nature (London)* **275**, 505–510.
10. Wilson, J. T., Forget, B. G., Wilson, L. B. & Weissman, S. M. (1977) *Science* **196**, 200–202.
11. Wilson, J. T., Wilson, L. B., de Riel, J. K., Villa-Komaroff, L., Efstratiadis, A., Forget, B. G. & Weissman, S. M. (1978) *Nucleic Acids Res.* **5**, 563–581.
12. Curtiss, R., III, Pereira, D. A., Hsu, J. C., Hull, S. C., Clarke, J. E., Maturin, L. J., Sr., Goldschmidt, R., Moody, R., Inoue, M. & Alexander, L. (1977) in *Recombinant Molecules: Impact on Science and Society*, Proceedings of the 10th Miles International Symposium, eds. Beers, R. F., Jr. & Bassett, E. G. (Raven, New York), pp. 46–56.
13. National Institutes of Health (1976) *Fed. Reg.* **41**, 27902–27943.
14. Huisman, T. H. J. & Dozy, A. M. (1965) *J. Chromatogr.* **19**, 160–169.
15. Efremov, G. D., Huisman, T. H. J., Smith, L. L., Wilson, J. B., Kitchens, J. L., Wrightstone, R. N. & Adams, H. R. (1969) *J. Biol. Chem.* **244**, 6105–6116.
16. Garver, F. A., Jones, C., Baker, M. M., Altay, G., Barton, B. P., Gravely, M. & Huisman, T. H. J. (1976) *Am. J. Hematol.* **1**, 459–469.
17. Garver, F. A., Baker, M. M., Jones, C. S., Gravely, M., Altay, G. & Huisman, T. H. J. (1977) *Science* **196**, 1334–1336.
18. Garver, F. A., Altay, G., Baker, M. M., Gravely, M. & Huisman, T. H. J. (1978) *J. Immunol.* **121**, 1297–1301.
19. Altay, G., Garver, F., Bannister, W. H., Grech, J. L., Felice, A. & Huisman, T. H. J. (1977) *Biochem. Genet.* **15**, 915–923.
20. Garver, F. A. & Talmage, D. W. (1975) *J. Immunol. Methods* **7**, 271–282.
21. Maniatis, T., Kee, S. G., Efstratiadis, A. & Kafatos, F. C. (1976) *Cell* **8**, 163–182.
22. Reichlin, M., Bucci, E., Fronticelli, C., Wyman, J., Ioppolo, E. A. C. & Rossi-Fanelli, A. (1966) *J. Mol. Biol.* **17**, 18–28.
23. Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K. & Riggs, A. D. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 106–110.

TAB Y

Proc. Natl. Acad. Sci. USA
Vol. 76, No. 11, pp. 5596–5600, November 1979
Biochemistry

Synthesis of simian virus 40 t antigen in *Escherichia coli*

(recombinant plasmid/hybrid ribosome binding site/eukaryotic gene expression/amino acid sequence analysis)

THOMAS M. ROBERTS*, ILAN BIKEL†, R. ROGERS YOCUM*, DAVID M. LIVINGSTON†, AND MARK PTASHNE*

*The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138; and †Sidney Farber Cancer Institute and the Departments of Medicine, Peter Bent Brigham Hospital and Harvard Medical School, Boston, Massachusetts 02115

Contributed by Mark Ptashne, August 22, 1979

ABSTRACT **Plasmids are constructed by using recombination *in vitro* according to Roberts, T. M., Kacich, R. & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. USA* 76, 760–764 in which the t antigen gene of simian virus 40 is fused to a promoter of the *Escherichia coli lac* operon. In the fusions, transcription commences at the *lac* promoter, and, in some of the fusions, translation begins at the ATG initiator codon of the *t* gene. This translation is directed most efficiently by those plasmids in which the *lac* sequences abut the *t* gene such that a hybrid ribosome binding is encoded. In this case, the Shine–Dalgarno sequence is of *lac* origin but the ATG derives from the *t* gene. Translation from this initiator codon is greatly decreased if the *lac* sequences are separated from the ATG by 17 base pairs and is abolished if the AT of this triplet is deleted. Cells bearing the productive fusions synthesize a 20,000-dalton protein with t antigenic determinants. This protein has an isoelectric point(s) indistinguishable from that of t antigen isolated from simian virus 40-transformed cells. Moreover, a partial sequence of the amino-terminal region of the bacterial product is that predicted for authentic t antigen. We conclude that these bacteria are producing a protein, the sequence of which is identical to that of authentic t antigen unfused to other polypeptides.**

The T antigens of simian virus 40 (SV40) are virus-encoded proteins that play a central role in the process of virus-induced neoplastic transformation (1–8). One of these proteins, the t antigen (t), is a 174-amino acid polypeptide whose coding sequence is known. Unlike the related protein SV40 T antigen (T), it is cytoplasmic and does not bind to DNA *in vitro* (ref. 1; J. Griffin and D. Livingston, unpublished results). Analysis of t function *in vitro* has been limited by the fact that it is difficult to purify from SV40-infected cells. The DNA encoding t has no intervening sequences and, if supplied the appropriate *cis*-acting control sequences, its expression in a bacterium should produce a polypeptide identical in sequence to t.

In this paper, we describe the construction of plasmids that direct the synthesis of apparently authentic SV40 t in bacterial cells. The essential feature of the construction is the fusion of the *t* gene to an *Escherichia coli* DNA fragment bearing a promoter and a so-called Shine–Dalgarno (SD) sequence (9–11), both from the *lac* operon. The former insures efficient transcription of the *t* gene. The latter, when positioned at the appropriate distance from the ATG signaling the translational start, directs binding of the mRNA to ribosomes and proper initiation of translation. We refer to such a regulatory sequence as a "hybrid ribosome binding site". The product of translation is not a bacterial–eukaryotic fusion protein. Rather, our experiments strongly suggest that, as expected, it has the primary structure of t. The approach we use to position the DNA fragment containing the bacterial control sequences is based on the principle and methods utilized to maximize the production of two regulatory proteins encoded by phage λ, repressor and *cro* protein (12, 13).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Animal Cells and Viruses. The SV80 strain of SV40-transformed human fibroblasts (14) served as a source of authentic t. SV40, strain 777, was grown and its DNA was isolated as described (15). Protein was labeled *in vivo* with [^{35}S]methionine (specific activity, 400–600 Ci/mmol; 1 Ci = 3.7 × 10^{10} becquerels) at a concentration of 50–100 μCi/ml as described (16).

Bacterial Cells and Radiolabeling of Protein. *E. coli* strains MM294 (*endo I*$^-$, *B*$^-$, r_k^-, m_k^+) and RB113·(*recA13, uvrA6, thi-1, thr1, leuB6, proA2, argE3, his-4, mtl-1, xyl-5, tsx33, strA31, gal2, ara14, lacY1*) were used. Labeling of total cell protein was carried out in MM294 by using $Na^{35}SO_4$ under the condition of Roberts and Roberts (17). Plasmid-coded proteins were labeled in RB113 by using [^{35}S]methionine under the conditions of Sancar *et al.* (18) as modified by R. Brent (unpublished), the 'maxicell technique'. This technique specifically labels with radioisotope plasmid-coded protein. A typical experiment is performed by lightly UV irradiating *recA*$^-$ *uvrA*$^-$ cells bearing the plasmid of interest. The cells are then incubated in nutrient medium for 3 hr at which time cycloserine is added to 100 μg/ml. The incubation is then continued for 10–14 hr during which time most of the host chromosomal (but not plasmid) DNA is destroyed. Finally, the cells are washed in sulfur-free medium and incubated in the presence of [^{35}S]-methionine for another 2 hr. For details, see Sancar *et al.* (18).

Plasmid Construction and Analysis. Enzymes and techniques for plasmid construction and analysis have been reported (12, 13, 19). All plasmids were constructed and propagated under conditions conforming to the standards outlined in the National Institutes of Health Guidelines for Recombinant DNA Research.

T Antigen Immunoprecipitation and Gel Electrophoresis. Total cell protein from SV80 cells was labeled with [^{35}S]methionine, extracted, and immunoprecipitated either with serum from hamsters bearing SV40-induced tumors or with serum from tumor-free animals as described (16). For the immunoprecipitation of t, polypeptides from *E. coli* extracts were prepared and immunoprecipitated in the presence of an excess of unlabeled *E. coli* 294 extract as described by Roberts and Roberts (17). Sodium dodecyl sulfate/polyacrylamide gel electrophoresis was as described by Laemmli (20). After fixation, staining, and destaining (where indicated), gels were

Abbreviations: SV40, simian virus 40; kd, kilodaltons; SD, Shine–Dalgarno; t, t antigen; T, T antigen.

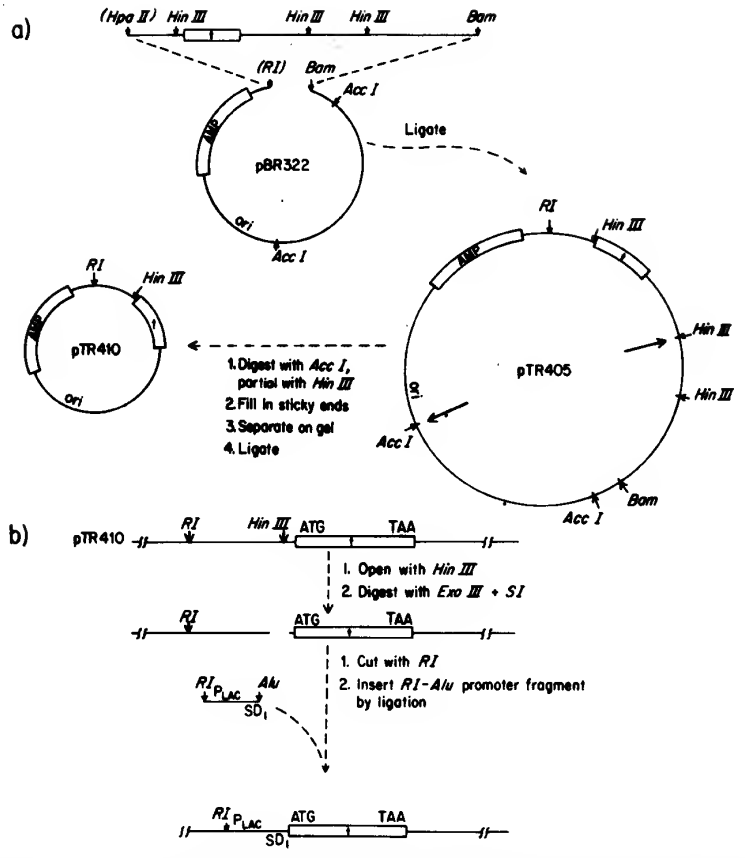


FIG. 1. Schematic representation of plasmid construction. The approximate location of several restriction endonuclease cleavage sites are shown for a large fragment of the plasmid pBR322, for an SV40 DNA fragment bearing the *t* gene, and for a DNA fragment bearing the promoter (P_{lac}) and SD sequence of the *lac*Z gene (SD_1). [See Backman and Ptashne (12) for the source of this fragment.] The position of the origin of replication (*ori*) and β-lactamase gene of pBR322 (*AMP*) are indicated. ATG and TAA are the start and stop signals for translation of *t*. Construction steps were as follows: (*i*) the *t* gene was cloned in pBR322. SV40 DNA was cleaved with *Hpa* II and the resulting sticky ends were rendered flush with DNA polymerase I and the four deoxynucleotide triphosphates (these "filled-in" restriction sites are designated with parentheses). Subsequent cleavage with *Bam* generated a 3000-base pair fragment carrying the *t* gene. This fragment was joined to a pBR322 backbone generated by R_I cleavage, treatment with polymerase I to fill in the R_I end, and finally *Bam* cleavage. In the resulting plasmid, pTR405, the filled-in R_I site of pBR322 has been fused to the filled-in *Hpa* II site on SV40 to regenerate an R_I site. (*ii*) A portion of pTR405 containing two unwanted *Hin*dIII sites was deleted: pTR405 DNA was digested to completion with *Acc* I and partially digested with *Hin*dIII. The resulting sticky ends were rendered flush and the various digestion products were separated by agarose gel electrophoresis. The correct partial digestion product (the ends of which are designated by long arrows inside the pTR405 circle) was isolated from the gel, recircularized with T4 DNA ligase, and used to transform *E. coli*. In the resulting plasmid, pTR410, the first *Hin*dIII site downstream from the *t* gene has been fused to the *Acc* I site near the origin of pBR322 replication. Thus, pTR410 has only one *Hin*dIII site located just upstream from the translational start of t. In *b*, plasmid pTR410 and derivatives of it are represented linearly. (*iii*) pTR410 was opened with *Hin*dIII, and varying amounts of DNA were removed with *Exo* III and S1 nuclease. (*iv*) The resected plasmids were cut with R_I, and the DNA fragment bearing the promoter and SD sequence of the *lac* operon was inserted. The product is a series of plasmids bearing the *lac* promoter and its associated SD sequence at various distances from the SV40 t sequence.

usually dried and autoradiographed on Kodak XR-1 film. On occasion, gels were impregnated with 2,5-diphenyloxazole and fluorographed by the method of Bonner and Laskey (21).

Two-Dimensional Gel Electrophoresis. Two-dimensional gel electrophoresis was performed by the nonequilibrium technique, essentially as described by Crawford and O'Farrell (22). A 12.5% polyacrylamide slab running gel was utilized in the second dimension. Gels were prepared for fluorography, dried, and exposed to x-ray film as described above.

Automated Amino Acid Sequence Analysis. The 20 kilodalton (kd) protein was labeled with [^{35}S]methionine in cells carrying pTR436 by using the "maxicell" technique (18). It was additionally purified by immunoprecipitation and sodium dodecyl sulfate gel electrophoresis. The material was eluted from the gel, and 50,000 cpm were analyzed in a Beckman 890B Automated Sequenator by using a 0.1 M Quadrol program (23). The radioactivity produced at each cycle was measured in a liquid scintillation counter. Unlabeled apomyoglobin (700 μg) was added to the initial sample as an internal standard.

RESULTS

Plasmid Construction, Screening, and Initial Characterization. Fig. 1 shows the method used to construct plasmids that direct synthesis of t in bacteria. The essential steps were: (*i*) a SV40 DNA fragment containing the *t* gene was cloned in pBR322 to produce pTR405. (*ii*) To delete unwanted restriction sites, pTR405 was reduced in size by restriction and ligation to produce pTR410. The important features of this plasmid are

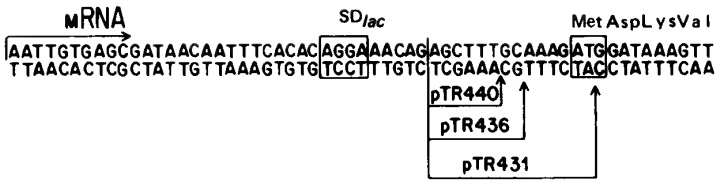


FIG. 2. Sequences around beginning of *t* gene in a series of fusion plasmids. Shown is a portion of the sequence of pTR422 extending from the startpoint of transcription of the *lac* promoter on the left (straight arrow) to the startpoint of translation of the *t* gene on the right. Also shown is the SD sequence of the *lac* operon. All plasmids produced by the *Exo* III/S_1 method may be considered deletions of pTR422. The extent of these deletions in three of the plasmids is indicated by brackets.

a unique *Hin*dIII site just (14 base pairs) upstream from the start point of translation of the *t* gene and a unique *Eco*R1 site further (400 base pairs) upstream. (*iii*) pTR410 DNA was cleaved with *Hin*dIII and shortened to varying extents with exonuclease III and S1 nuclease. In this way, varying amounts of the DNA between the *Hin*dIII site and the start-point of translation of t were removed. (*iv*) The resected plasmid was cut with *Eco*R1, and a DNA fragment bearing the prokaryotic control signals was inserted into the plasmid backbone. Because this fragment was bounded by *Eco*R1 and *Alu* ends, its direction of insertion was fixed.

Plasmids constructed as described in Fig. 1 were used to transform *E. coli*, and ampicillin-resistant clones were screened for production of t by immune precipitation. Approximately half the clones produced a protein of approximately 20 kd which was specifically precipitated by anti-T sera. The nucleotide sequence of about 80 base pairs spanning the fusion of bacterial and *t* gene sequences was determined for plasmids isolated from several of the clones, both positive and negative for immunoprecipitable 20-kd protein.

In a separate experiment, the enzymatic resection step was omitted, and the sticky ends produced by *Hin*dIII digestion of pTR410 DNA were rendered flush with DNA polymerase I and the four deoxyribonucleotide triphosphates before insertion of the fragment bearing the *lac* control sequences. The resulting plasmid was pTR422.

Fig. 2 shows sequences from three of the resulting plasmids as well as the unresected pTR422. In pTR440 and pTR436, the distances separating the SD sequence of the *lac* operon are 11 and 9 base pairs, respectively. In pTR431, the first two bases of the small *t* gene ATG were removed.

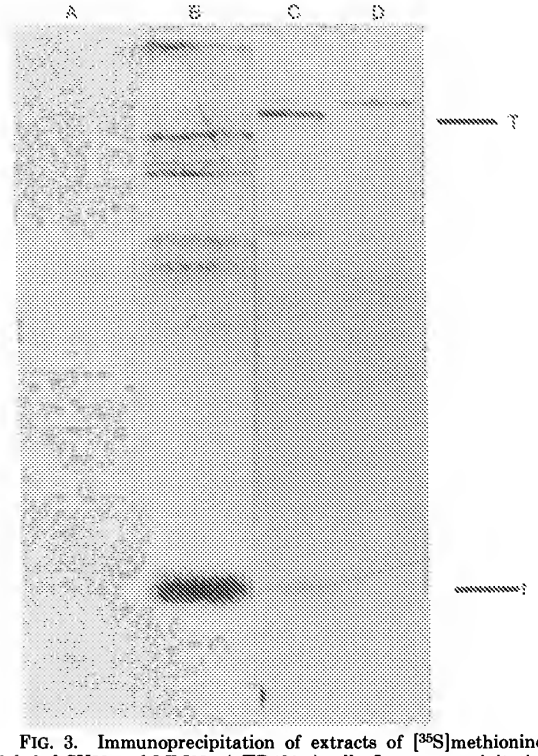


FIG. 3. Immunoprecipitation of extracts of [^{35}S]methionine-labeled SV80 and MM294 (pTR421$_{36}$) cells. Immunoprecipitation was performed on two identical aliquots of each labeled cell extract in parallel as described (16, 17). In one instance, 10 μl of hamster anti-T serum was employed, and, in the other case, 10 μl of control hamster serum was added. Immunoprecipitates were dissolved and electrophoresed in 12% sodium dodecyl sulfate/polyacrylamide gels with 4% polyacrylamide stacking gels, and the gels were autoradiographed. Lanes: A, control serum precipitate of MM294 (pTR436) cell extract; B, anti-T serum precipitate of the same extract; C, anti-T serum precipitate of SV80 cell extract; D, control serum precipitate of the same extract.

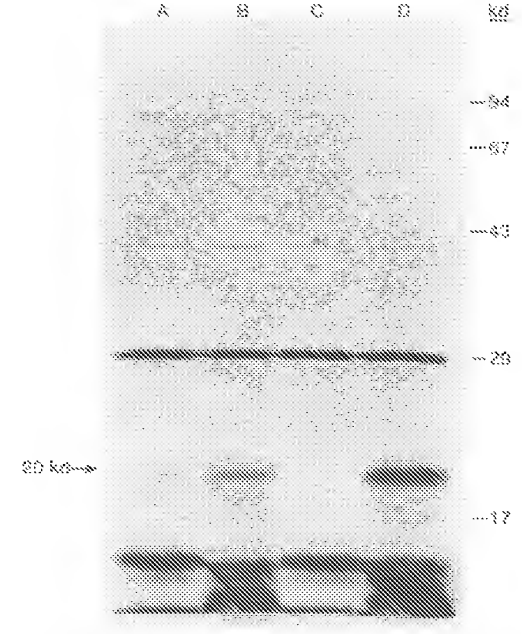


FIG. 4. Synthesis of t directed by various fusion plasmids. The maxicell technique was used to specifically label protein whose synthesis is directed by four fusion plasmids. After the cells were labeled, they were disrupted and the contents were examined directly by polyacrylamide gel electrophoresis and autoradiography. The position of a 29-kd protein (presumably β-lactamase) and of a 20-kd species (presumably t) are indicated. The plasmids used for the experiments are as follows; lanes: A, pTR422; B, pTR440; C, pTR431; D, pTR 436. Approximately 4×10^6 cpm of each extract was applied to the gel.

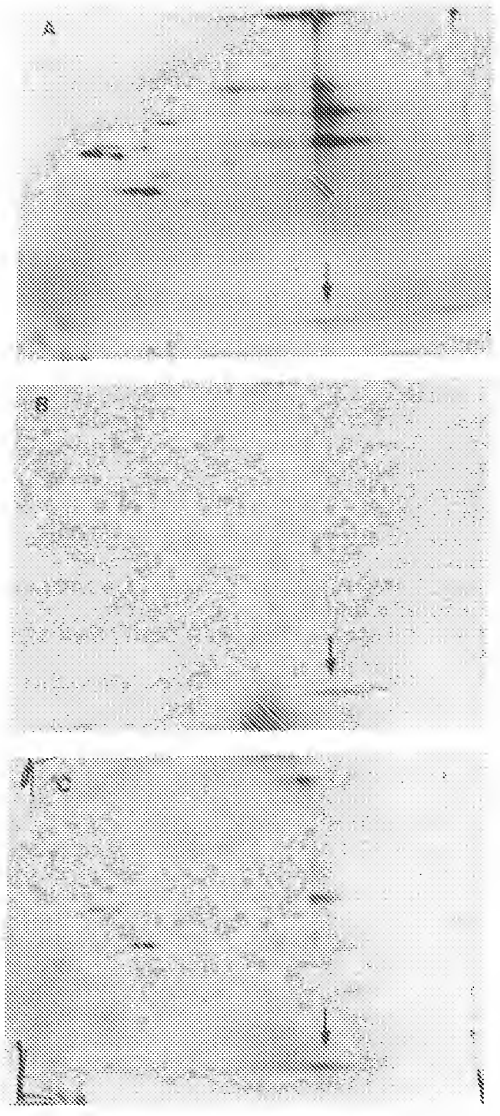


FIG. 5. Two-dimensional gel electrophoresis of bacterial and mammalian T antigens. T and t, labeled with [^{35}S]methionine, were isolated from an extract of SV80 cells by immunoprecipitation. The 20-kd protein was prepared from RB113 (pTR436) labeled with [^{35}S]methionine by the maxicell technique. It was purified further by immunoprecipitation. Proteins were eluted, reduced, and alkylated with *N*-ethylmaleimide and subjected to nonequilibrium, two-dimensional gel electrophoresis (22). Separation in the vertical dimension is by size and in the horizontal, by charge. (*A*) SV80 immunoprecipitate; 70,000 cpm was applied. (*B*) RB113 (pTR436) immunoprecipitate; 20,000 cpm was applied. (*C*) Mixture of SV80 immunoprecipitate (30,000 cpm) and RB113 (pTR436) immunoprecipitate (10,000 cpm). The gels were fluorographed and exposed for 7 days. One possible source of the t streaking is heterogeneous alkylation. The arrow denotes the 24-kd protein. The two prominent species in the left part of *A* and *C* are tubulin (56 kd) and actin (43 kd).

Characterization of Plasmid-Encoded Protein. Fig. 3 shows that treatment of ^{35}S-labeled total cell protein from clone MM294 (pTR436) with anti-T serum precipitates a protein of the same mobility in a sodium dodecyl sulfate/polyacrylamide gel as t (20 kd). The figure also shows that control hamster serum fails to precipitate this protein.

Plasmid-encoded proteins may be specifically radiolabeled by the maxicell technique of Sancar *et al.* (18). The experiment of Fig. 4 utilized this technique to show that pTR436 and pTR440 each direct the synthesis of two proteins. One was of the size expected for β-lactamase (29 kd) and the other was the same size as t (20 kd). This latter species was also specifically precipitated by anti-T serum (not shown). Both of these plasmids produced comparable amounts of the 29-kd protein, but pTR436 directed synthesis of almost twice as much of the 20-kd species as did pTR440. In the former, the SD sequence was 9 base pairs from the AUG of t, and, in the latter, it was 11 base pairs. In striking contrast are results with strains bearing pTR431 and pTR422. In both cases, β-lactamase was produced but synthesis of the 20-kd protein was absent in the former and barely detectable in the latter. These results show that deletion of the A and T of the initiating ATG of t abolishes synthesis of the protein and that positioning of the *lac* SD sequence is crucial. Fig. 4 also shows that smaller species (≤10 kd) are also produced by our plasmids. In those cases (lanes B and D) in which t was produced, we suspect that at least some of these smaller species are breakdown products of that protein. In the cases (lanes A and C) in which little or no t was synthesized, a relatively predominant species of about 10 kd appeared, which we suspect is initiated at an internal ATG in the *T* gene.

Comparative Two-Dimensional Gel Electrophoresis of t Antigen and the *E. coli* 20-kd Anti-T-Reactive Polypeptide. As assayed by two-dimensional gel electrophoresis, authentic t was indistinguishable from the presumptive *t* gene product made in *E. coli* (Fig. 5). Both formed streaks during isoelectric focusing and subsequently migrated as ≈24-kd species in the second dimension. Crawford and O'Farrell (22) have shown that the apparent molecular weight of alkylated t is approximately 24 kd. Protein streaking is apparently not an artifact of the gel system because neither tubulin nor actin, also observed in these immune precipitates, streaked. These two proteins have been found to be nonspecifically precipitated by such sera under these conditions (22). Control serum precipitates of the SV80 extract revealed only tubulin and actin spots whereas those of pTR436 revealed no spots (data not shown).

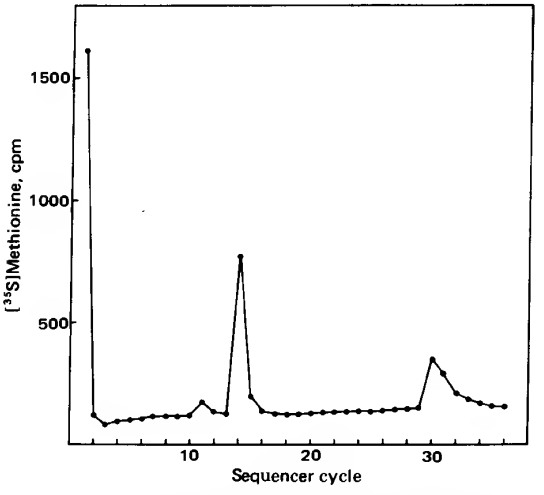


FIG. 6. Location of methionines in the amino-terminal segment of bacterial t antigen. The 20-kd protein labeled with [^{35}S]methionine (50,000 cpm) was analyzed by automated Edman degradation. The product of each cycle of degradation was quantitated in a liquid scintillation counter.

Partial Amino-Terminal Sequence of the 20-kd Protein Product. Fig. 6 shows that, of the first 36 residues of the 20-kd protein isolated from bacteria, methionine is present only at positions 1, 14, and 30. This is known to be the case for authentic t (24). The analysis was performed by automated Edman degradation of bacterial t labeled with [^{35}S]methionine. As is often the case in such analyses, the yield of radioactive amino acid was 25–40% of the amount applied. Thus this result, in addition to partially confirming the expected sequence of the bacterial product, suggests that a significant portion, if not all, of the molecules have an unblocked amino terminus.

DISCUSSION

Our plasmid-bearing cells produce authentic t by the following criteria: (*i*) a plasmid-encoded 20-kd protein is specifically immunoprecipitated with anti-T serum; (*ii*) the protein has precisely the mobility of t in both one- and two-dimensional gels; (*iii*) the position of methionine residues (1, 14, and 30) at the amino terminus of the protein is that of t; (*iv*) synthesis is abolished if the first two bases of the t initiation codon are deleted, and synthesis is barely detectable if the distance between the *lac* SD sequence and the ATG of t is large (17 base pairs). The latter result also supports the notion that formation of a hybrid ribosome-binding site bearing appropriately positioned SD and ATG sequences is essential to translation of *t*. We assume, without explicit proof, that translation of *t* in our bacteria stops at the first in-phase termination codon. This UAA, which immediately follows the codon for amino acid 174, is presumed to be the termination signal for authentic *t* (25–27).

The sequenator results show that at least a fraction of the protein bears an unmodified amino-terminal methionine, but we cannot exclude the possibility that some of the protein molecules have a modified amino terminus. t isolated from animal cells bears an acetylated amino-terminal methionine (24). It remains to be seen whether this secondary modification will affect the protein's activity.

We are not certain of the level of t produced in our bacterial cells. Indirect estimates suggest that our best fusions produce roughly 1000–5000 monomers per cell. We have not systematically varied the distance between the *lac* SD sequence and the ATG of t nor the sequence of the intervening bases. Such modifications might produce a strain that will yield significantly greater quantities of protein.

Most of the reports of synthesis of higher eukaryotic proteins in bacteria have described the production of fusion proteins composed of both prokaryotic and eukaryotic sequences (28, 29). In one case, it was suggested that the protein might be initiating at the correct ATG, but this was not directly demonstrated (30). Our experiments show that a message bearing a hybrid ribosome-binding site—i.e., sequences derived partly from the bacterium (the SD sequence) and partly from the eukaryotic gene (the ATG, etc.)—can be correctly translated into protein. This provides a rational approach to the problem of obtaining expression of eukaryotic genes in bacteria.

We thank Daniela Faibes for expert assistance with a number of the experiments and Dr. Lan Bo Chen for help with two-dimensional electrophoresis. This work was supported by grants from the National Science Foundation and the National Institutes of Health to D.M.S. and M.P.

1. Prives, C., Gilboa, E., Revel, M. & Winocour, E. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 457–461.
2. Crawford, L. V., Cole, C. N., Smith, A. E., Paucha, E., Tegtmeyer, P., Rundell, K. & Berg, P. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 117–121.
3. Martin, R. & Chou, J. Y. (1975) *J. Virol.* **15,** 599–612.
4. Osborn, M. J. & Weber, K. (1975) *J. Virol.* **15,** 636–644.
5. Tegtmeyer, P. (1975) *J. Virol.* **15,** 613–618.
6. Bouck, N., Beales, N., Shenk, T., Berg, P. & di Mayorca, G. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2473–2477.
7. Sleigh, M. J., Topp, W. C., Hanich, R. & Sambrook, J. F. (1978) *Cell* **14,** 79–88.
8. Martin, R. G., Setlow, V. P., Edwards, C. A. F. & Vembu, D. (1979) *Cell* **17,** 635–643.
9. Shine, J. & Dalgarno, L. (1975) *Nature (London)* **254,** 34–38.
10. Steitz, J. A. & Jakes, K. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 4734–4738.
11. Steitz, J. A. (1978) in *Biological Regulation and Development,* ed. Goldberger, R. (Plenum, New York), Vol. 1, pp. 349–399.
12. Backman, K. & Ptashne, M. (1978) *Cell* **13,** 65–71.
13. Roberts, T. M., Kacich, R. & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 760–764.
14. Todaro, G., Green, H. & Swift, M. R. (1966) *Science* **153,** 1252–1254.
15. Jessel, D., Hudson, J., Landau, T., Tenen, D. & Livingston, D. (1976) *Cell* **8,** 535–545.
16. Griffin, J., Light, S. & Livingston, D. (1978) *J. Virol.* **27,** 218–226.
17. Roberts, J. W. & Roberts, C. W. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 147–151.
18. Sancar, A., Hack, A. M. & Rupp, W. D. (1979) *J. Bacteriol.* **137,** 692–693.
19. Roberts, T. M. & Lauer, G. D. (1979) *Methods Enzymol,* in press.
20. Laemmli, U. (1970) *Nature (London)* **227,** 680–685.
21. Bonner, W. M. & Laskey, R. A. (1974). *Eur. J. Biochem.* **46,** 83–88.
22. Crawford, L. V. & O'Farrell, P. Z. (1979) *J. Virol.* **29,** 587–596.
23. Bauer, A. W., Margolies, M. N. & Haber, E. (1975) *Biochemistry* **14,** 3029–3035.
24. Paucha, E., Mellor, A., Harvey, R., Smith, A. E., Hewick, R. W. & Waterfield, M. B. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2165–2169.
25. Reddy, V. B., Ghosh, P. K., Lebowitz, P., Tiatak, M. & Weissman, S. M. (1979) *J. Virol.* **30,** 279–296.
26. Reddy, V. B., Thimmappaya, B., Dhar, R., Subramanian, K. N., Zain, B. S., Pan, J., Ghosh, P. K., Celma, M. L. & Weissman, S. M. (1978) *Science* **200,** 494–502.
27. Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van De Voorde, A., Van Heuverswyn, H., Van Herrweghe, J., Volckaert, G. & Ysebaert, M. (1978) *Nature* **273,** 113–120.
28. Mercereau-Puijalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, F. & Kourilsky, P. (1978) *Nature (London)* **275,** 505–510.
29. Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3727–3731.
30. Chang, A. C. Y., Nunberg, J. H., Kaufman, R. J., Erlich, H. A., Schimke, R. T. & Cohen, S. N. (1978) *Nature (London)* **275,** 617–624.

TAB Z

Nature Vol. 283 10 January 1980

Influenza antigenic determinants are expressed from haemagglutinin genes cloned in *Escherichia coli*

J. S. Emtage, W. C. A. Tacon, G. H. Catlin, B. Jenkins, A. G. Porter & N. H. Carey

Searle Research Laboratories, Lane End Road, High Wycombe, Bucks, UK

A gene sequence for the fowl plague virus (FPV) haemagglutinin molecule has been inserted into a bacterial plasmid such that its transcription is under the control of a promoter derived from the tryptophan operon. Such plasmids direct the synthesis of a protein that reacts specifically with antisera to FPV haemagglutinin. Evidence is also presented that in some cases DNA inserted at the HindIII *site of pBR322 is expressed.*

INFLUENZA is still a major disease of man. It causes death primarily in the elderly, many of whom have chronic heart or vascular disease and other disorders. Severe infections of healthy individuals also occurs but in these cases morbidity, leading to incapacity to work, rather than mortality usually results. The morbidity figures however are of great economic importance. For example, the Hong Kong outbreak of 1968–70 resulted in 25 million lost working days in England[1], while in the US the effect of the Hong Kong variant of 1968–69 was costed at 3.8 billion dollars[2]. Thus, prevention of influenza epidemics and pandemics would be of value both economically and socially.

The main structure involved in immunity against influenza is the haemagglutinin (HA) surface glycoprotein[3,4]. The functional HA subunit, one of the spikes on the virus surface, is a triangular rod-shaped glycoprotein with a molecular weight of ~250,000 and is comprised of three HA monomers[5]. The HA monomers are synthesised as single polypeptide chains containing, in the case of FPV, an 18-amino acid precursor peptide at the N-terminus[6]. During maturation and virus assembly the pre-HA is further processed to HA1 and HA2 which remain linked by disulphide bridges[7,8].

Recent advances in genetic engineering and our knowledge of the structure and sequence of bacterial operons and control elements now allow the construction of new bacterial strains with the potential of synthesising large quantities of viral proteins (antigens) which may ultimately be useful as vaccines. In this article we describe the controlled production of haemagglutinin antigen from an HA gene cloned in *Escherichia coli*; this is the first step towards testing the feasibility of large scale antigen production by these means.

Experimental design

We have previously cloned at the *Hin*dIII site of pBR322 a DNA fragment containing the control region of the *E. coli* tryptophan operon, and coding for the ribosome binding site and first seven amino acids of *trpE*. From this, we have produced a series of vectors, pWT111, pWT121 and pWT131, capable of ensuring that any inserted DNA is read in the correct phase[9].

We recently cloned and sequenced the gene for the FPV haemagglutinin protein[6]. The gene was cloned using *Hin*dIII linkers and can therefore be transferred between vectors. From the nucleotide sequences of the FPV HA gene and the *Hin*dIII sites of the pWT series, it was clear that, if inserted at the *Hin*dIII site of pWT121 in the correct orientation, the HA gene would be translated by readthrough from the *trpE* fragment (Fig. 1*A*).

The protein resulting from initiation at the *trpE* AUG would be a hybrid consisting of the following fragments in order: (1) an N-terminus of 7 amino acids from anthranilate synthetase; (2) 6 amino acids specified by linker DNA; (3) 6 phenylalanine residues from the $(T)_{19}$ region of the FPV cloned DNA; (4) 7 amino acids from the normally non-translated 5′-portion of the HA gene; (5) 558 amino acids comprising the haemagglutinin protein and its prepeptide[6] and, finally; (6) 5 amino acids specified by *Hin*dIII linker at the C-terminus[9]. This is a total of 589 amino acids with a total MW of 69,000.

It was also clear that the HA gene should not be expressed from the initiator AUG of the *trpE* fragment in pWT111 or pWT131. Recognition of the ribosome binding site and initiator AUG of the HA was not anticipated. Similarly, we did not expect expression of the HA gene inserted at the *Hin*dIII site of pBR322 in either orientation.

Construction of expression plasmids

pWT121 was restricted with *Hin*dIII, treated with alkaline phosphatase, ligated to the purified HA gene and the mixture used to transform *E. coli* K12 HB101. A total of 47 ampicillin resistant transformants was obtained (pWT121 in the absence of the HA gene produced 13 colonies). Of these 47 colonies, 19 were tetracycline sensitive (Tc^S) and the remainder tetracycline resistant (Tc^R). We selected 12 colonies (3Tc^S and 9 Tc^R) for further characterisation. Analysis of plasmid DNA by gel electrophoresis after *Hin*dIII restriction showed that 5 of the 9 Tc^R colonies contained plasmids with a DNA insert. The other four contained the parent plasmid.

The orientation of the inserted DNA was determined by restriction enzyme analysis of the plasmids. We denote the two orientations as either R or L and define R and L orientation in terms of the direction of transcription required for HA

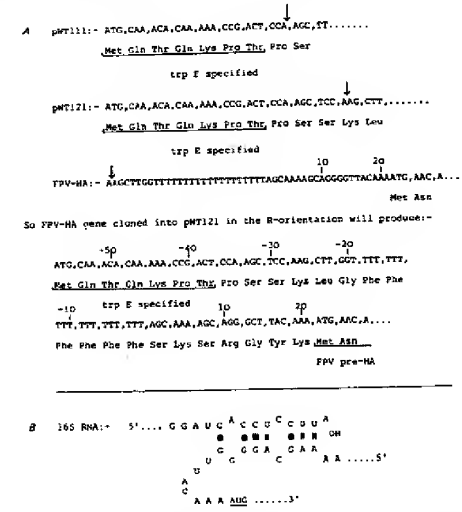


Fig. 1 *A*, Predicted nucleotide sequences around the *Hin*dIII sites of pWT111, pWT121 and the FPV-HA gene. Arrows indicate the position of the *Hin*dIII sites and, for clarity, only the sequence of the coding strand is shown. The protein sequence of the N-terminus of anthranilate synthetase has previously been reported by Lee *et al*[10]. The numbering system is bidirectional beginning with the 5′-nucleotide of the gene 4 complementary RNA. *B*, Potential base pairing between the 3′oligonucleotide of *E. coli* 16S RNA[11] and the pre-AUG region of the haemagglutinin mRNA.

0028-0836/80/020171—04$01.00

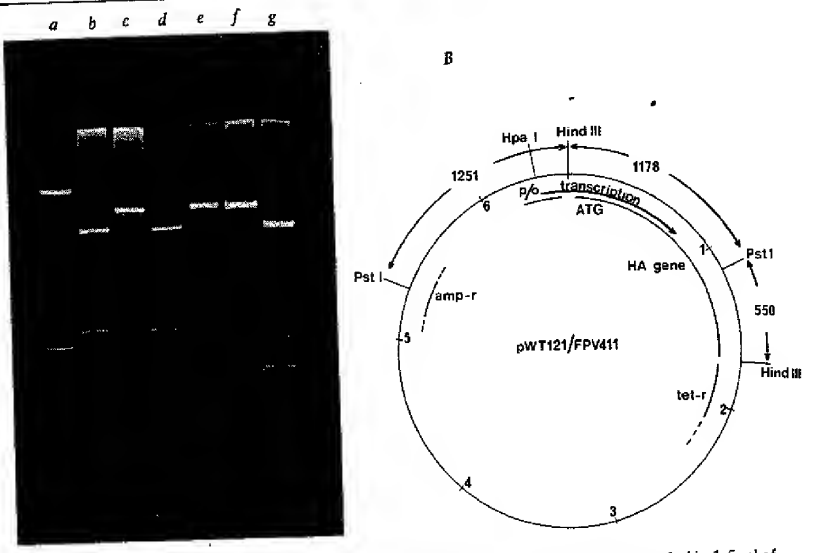


Fig. 2 *A*, Orientation of the FPV gene in pWT111, pWT121 and pBR322. pBR322, pWT111 or pWT121 (10 μg) was limit digested with *Hin*dIII. The resulting 5′-terminal phosphates of the vectors were removed by treatment with 20 μg bacterial alkaline phosphatase for 30 min at 37 °C in a 25 μl incubation containing 20 mM Tris-HCl *p*H 7.5 and 0.1% SDS. After phenol extraction and ethanol precipitation 0.2 μg of each DNA was ligated to ~40 ng FPV-HA gene in a 20 μl incubation containing 50 mM Tris-HCl *p*H 7.5, 10 mM $MgCl_2$, 20 mM dithiothreitol, 1 mM ATP and 0.02 units T4 DNA ligase (New England Biolabs). After overnight incubation at 15 °C the mixtures were diluted to 100 μl with TCM (10 mM Tris-HCl *p*H 7.5, 10 mM $CaCl_2$, 10 mM $MgCl_2$) and used to transfect 200 μl $CaCl_2$-treated *E. coli* K12 HB101 by a previously described procedure[12]. Transformants present after growth for 16 h at 37 °C on L-agar[13], plus 100 μg ml^{-1} carbenicillin (Pyopen) were tested for tetracycline resistance by picking onto agar plates containing M9 salts, glucose and casamino acids[13] plus carbenicillin and tetracycline (10 μg ml^{-1}). For plasmid isolation colonies were picked and 100-ml cultures grown in M9 salts, glucose, casamino acids medium supplemented with carbenicillin. At an A_{600} of ~0.6, choramphenicol was added to a final concentration of 150 μg ml^{-1} and incubation continued for 16 h at 37 °C. Cells were killed by the addition of diethylpyrocarbonate[14] to 0.4%, collected by centrifugation, washed with TE buffer (10 mM Tris-HCl *p*H 7.5, 1 mM EDTA) and suspended in 1.5 ml of iced 25% sucrose in 50 mM Tris-HCl *p*H 8. Cells were lysed by the addition of 0.5 ml lysozyme (10 mg ml^{-1}, 0.5 ml EDTA (0.5 M, *p*H 8) and 2.5 ml iced Triton solution (0.1% Triton X-100, 50 mM Tris-HCl *p*H 8, 50 mM EDTA) at 5-min intervals. The resulting lysates were cleared by spinning at 15,000*g* for 20 min, extracted with phenol and then chloroform and finally precipitated by the addition of 0.1 vol 3M NaAc and 0.54 vol isopropanol. After 30 min at −20 °C the DNA was pelleted, washed with 70% ethanol, dried and dissolved in TE. *Pst*I digests were carried out on 0.1 μg DNA in conditions recommended by the vendor (New England Biolabs) with the modification that RNase (50 μg ml^{-1}) was included in the incubation. Digests were fractionated on a 1.4% agarose slab gel[15] and bands visualised under UV light after staining with ethidium bromide. Lane *a* contains PM-2 DNA digested with *Hin*dIII. Bands seen are 5,400, 2,350 and 1,050 base pairs. The other lanes contain *Pst*I digests of the following plasmids: *b*, pWT121/FPV 411(R); *c*, pWT121/FPV 412(L); *d*, pWT111/FPV 502(R); *e*, pWT111/FPV 503(L); *f*, pBR322/FPV 604(L); *g*, pBR322/FPV 605(R). *B*, Structure of pWT121/FPV 411(R). pWT121 contains a single *Hin*dIII site downstream from the *trp* promoter[9]. The plasmid shown contains the FPV-HA gene cloned at the *Hin*dIII site in the R orientation. The distances between the *Pst*I and *Hin*dIII sites are shown as well as the direction of transcription from the *trp* promoter and the position of the initiator ATG. A full description of the construction of the pWT plasmid series has been published elsewhere[9]. Briefly, the 497-base pair *Hin*fI fragment containing the *trp* promoter/operator region as well as the nucleotides specifying the leader sequence and first seven amino acids to *trpE* (ref. 10) was isolated and cloned, using *Hin*dIII linkers, into the *Hin*dIII site of pBR322. From this a plasmid, pWT101, was isolated containing the *trp* promoter in the R orientation. In pWT101 tetracycline resistance is controlled from the *trp* promoter. Elimination of the *Hin*dIII site upstream of the *trp* promoter produced the vector pWT111 and phase changing at the *Hin*dIII site of pWT111 using DNA polymerase I and *Hin*dIII linkers produced pWT121 and pWT131.

production. That is, a gene in the R orientation required rightward (clockwise) transcription for expression; in the pWT plasmid series this is the orientation for expression from the *trp* promoter. For clarity we have included the letter (R) or (L) after the plasmid number to indicate its orientation. pWT121 contains a single *Pst*I site, in the ampicillin gene[16], 1,251 base pairs from the *Hin*dIII site and the FPV HA gene contains a *Pst*I site 1,178 base pairs from the 5′ end of the coding strand[6]. Thus *Pst*I digests of the recombinant plasmids should indicate which are correctly orientated with respect to the *trp* promoter. Representatives of the two FPV-gene containing groups are shown in Fig. 2*A* (*b* and *c*). The Tc^S plasmids gave rise to a band of 1,800 base pairs on *Pst*I digestion while the Tc^R plasmids produced one of 2,450 base pairs. These are consistent with the fragments expected from the plasmid shown in Fig. 2*B* and indicate that all the Tc^S plasmids contained inserts in the L orientation while the Tc^R plasmids contained inserts in the R orientation.

In a similar way we re-cloned the HA gene into the *Hin*dIII site of pWT111 and pBR322. Again the pWT111/FPV plasmids could be divided into Tc^R and Tc^S groups and *Pst*I restrictions (Fig. 2*A*, *d* and *e*) of representatives of these groups indicated that the HA gene was in the R orientation in the Tc^R group and in the L orientation in the Tc^S group. The pBR322/FPV plasmids however were all Tc^S but still contained L and R orientated genes as shown in Fig. 2*A* (*f* and *g*).

The correct phasing of the HA gene in pWT121/FPV 411(R) with the *trpE* AUG was confirmed by nucleotide sequencing. The HA gene contains an *Eco*RII restriction site at positions 36–40 (ref. 6). This site, however, is methylated and although not cleaved by *Eco*RII, it is cleaved by the *Eco*RII isoschizomer *Bst*N1. Therefore we digested pWT121/FPV 411(R) with *Bst*N1, labelled the 5′ ends with ^{32}P, recut the DNA with *Hpa*I and isolated the fragment containing the initiation region. The nucleotide sequence of this fragment was complementary to the sequence predicted in Fig. 1*A*, except for nucleotide −36 which was an A instead of a G. This was also the case for the corresponding position in pWT111. In both plasmids this position corresponds to the 3′-terminal nucleotide of the *Hin*dIII linker. The reason for the change is not clear; it might be due to random mutation or, more likely, to micro-heterogeneity or incomplete deblocking after chemical synthesis at the 3′ ends of the *Hin*dIII linker. In either event the effect in the mRNA strand is to have a serine codon (UCG) instead of a proline codon (CCG) but does not change the phase with respect to the *trpE* AUG.

Gene expression

E. coli colonies containing representatives of the plasmids described above were screened for FPV-HA antigen using a solid-phase immunological method[17]. Briefly, small cultures of individual colonies were grown, collected and lysed using lysozyme and Triton X-100. Any HA sequences present were

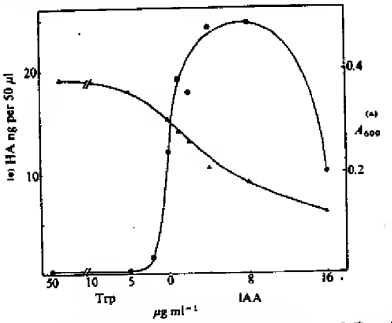


Fig. 3 Control of HA production. Overnight cultures of *E. coli* K12 HB 101 containing pWT121/FPV 411(R) were grown in M9 salts, glucose, casamino acids medium supplemented with carbenicillin (100 μg ml^{-1}). Aliquots of 100 μl were inoculated into 5 ml of the above medium containing either tryptophan or IAA as indicated. Cultures were grown for 2 h at 37 °C, the A_{600} determined and cells collected, lysed and assayed for HA content as described in Table 1.

Table 1 Radioimmunoassay of FPV-HA in lysates of bacteria containing influenza genes

Plasmid	Orientation of gene	Phenotype	HA content (ng per 50 μl)
pWT121/FPV 411	R	Tc^R	20.3
pWT121/FPV 412	L	Tc^S	5.0
pWT111/FPV 502	R	Tc^R	4.1
pWT111/FPV 503	L	Tc^S	2.1
pBR322/FPV 604	L	Tc^S	3.6
pBR322/FPV 605	R	Tc^S	0
pWT121	—	Tc^R	0
pBR322	—	Tc^R	0

Liquid cultures (5 ml) of individual colonies were grown in M9 medium. Cells were collected by centrifugation, and lysed, in a final volume of 1 ml, as described in Fig. 2A. This lysate was used for antigen assay without further treatment. HA-antigen assays were performed in polystyrene culture tubes (Nunc no. 1410) using an antibody sandwich technique[17]. Rabbit anti-FPV HA was purified by ammonium sulphate precipitation and DEAE-cellulose chromatography[18] before use and labelled with ^{125}I as described elsewhere[19]. Tubes were coated with 50 μl of 0.2M $NaHCO_3$ pH 9 containing 60 μg ml^{-1} purified, unlabelled IgG for 10 min at room temperature, washed with 3 × 1-ml aliquots of wash buffer (phosphate-buffered saline containing 0.1% bovine serum albumin, 0.1% NP-40 and 0.5% normal goat serum) and then incubated at room temperature either with known amounts (1–20 ng) of Sarkosyl-disrupted FPV or with the bacterial lysates. After 2 h, the tubes were washed with 4 × 1-ml aliquots of wash buffer and finally incubated at 4°C for 16 h with 50 μl wash buffer containing 200,000 c.p.m. ^{125}I-anti-HA. Tubes were again washed and counted in a Packard γ-counter. Antigen present was quantitated from standard curves relating the radioactivity bound to the amount of FPV present.

bound to a polystyrene tube coated with FPV-HA specific IgG. Bound antigen was then detected by incubating with high specific activity ^{125}I-anti FPV-HA. As indicated in Table 1, immune reactivity was detected from all colonies containing an FPV-HA gene inserted into the pWT plasmids and from pBR322 containing the HA gene in the L orientation. Neither of the parent plasmids nor pBR322 containing the HA gene in the R orientation produced any immune reacting material. Further, the positive reactions were abolished if the tubes were coated either with normal rabbit IgG or with anti A/Victoria—HA IgG instead of specific anti-FPV-HA IgG.

It was interesting to find expression of HA antigen from one orientation of pBR322/FPV. The *Hin*dIII site of pBR322 lies in the promoter region of the tetracycline gene and it is known that cloning at the *Hin*dIII site destroys this rightward transcribing promoter[20] unless the cloned DNA has its own promoter transcribing into the tetracycline gene(s)[21]. We therefore propose that a previously unknown promoter which transcribes in the leftward direction exists on the tetracycline gene side of the *Hin*dIII site of pBR322. It is necessary to postulate the existence of this promoter to explain the expression of the HA gene in pBR322/FPV 604(L), pWT121/FPV 412(L) and pWT111/FPV 503(L). The possibility that there is a pseudo-promoter sequence in the HA gene itself, for example the (T_{19}):(A_{19}) sequence, is ruled out as the plasmid pBR322/FPV 605(R) does not express its HA gene (Table 1).

This postulated promoter would explain the transcription of the HA gene in the L-orientated plasmids. How is translation explained? Inspection of the nucleotide sequence around the tetracycline–HA gene junction of pBR322/FPV 604(L) reveals nonsense triplets in all three translation phases[22]. Thus the bacterial translational system is probably recognising a nucleotide sequence on the HA mRNA itself and initiating protein synthesis at the AUG of the pre-peptide. Comparison of the sequence of the untranslated region of the HA gene with that of the 3′ end of prokaryotic 16S ribosomal RNA[11] reveals surprising complementarity (Fig. 1*B*). Consistent with this interpretation is the expression of HA antigen in pWT111/FPV 502(R). In this case the HA gene is in the wrong phase to be translated from the *trpE* AUG; protein synthesis initiating at the *trpE* AUG is terminated by the now in phase UGA triplet at position 23–25 in the HA gene. Thus we conclude that initiation is from the natural HA AUG at position 22–24 (Fig. 1*A*). Assuming equal transcription from the *trp* promoter in pWT121/FPV 411(R) and pWT111/FPV 502(R) and that the products have equal antigenic properties, then from the data of Table 1 we conclude that the eukaryotic ribosome binding site is recognised with an efficiency of 20% compared to the prokaryotic site.

The fivefold difference in HA expression between pWT121/FPV 411(R) and pWT111/FPV 502(R) could be due to differences in the plasmid complement of the respective cells. To exclude this possibility the relative copy numbers of the above plasmids were determined from agarose gels of lysed whole cells[9] and found to be the same (data not shown).

Hitherto the *Hin*dIII and *Eco*RI sites of pBR322 have been assumed to be non-expression sites because of their respective situations in and upstream of the tetracycline promoter[20,21]. This has been thought to be of value in assessing the containment categorisation when considering the safety of experiments in genetic engineering[23]. In view of the above findings it is clear that the validity of these assumptions should be reassessed. It seems likely that a potential ribosome binding site on the inserted gene will be of some importance.

Table 2 Control of HA expresion from the tryptophan promoter/operator region

Plasmid	Addition	HA* (ng per 50 μl)
pWT121/FPV 411(R)	Tryptophan	0.9
	IAA	59.1
pWT121/FPV 412(L)	Tryptophan	9.6
	IAA	4.9

Cultures (5 ml) of pWT121/FPV 411(R) and pWT121/FPV 412(L) were grown for 3 h at 37 °C in M9 salts, glucose, casamino acids medium supplemented with carbenicillin (100 μg ml^{-1}) and either IAA (20 μg ml^{-1}) or tryptophan (100 μg ml^{-1}). Cells were collected, lysed and assayed for HA content as described in Table 1.

* Results normalised to same A_{600}.

Control of gene expression

In vivo the tryptophan operon is controlled by the level of tryptophan[24] and control is best obtained in culture by using tryptophan as repressor or β-indole acrylic acid (IAA) as inducer. The latter is a competitive inhibitor that prevents tryptophan binding and thereby effectively inactivates the repressor[25].

We selected a representative colony from each of the two expressing groups (in pWT121) to examine HA synthesis in the above conditions. Our initial observation was that pWT121/FPV 411(R) grew very slowly in the presence of 20 μg ml^{-1} IAA; there was a threefold difference compared to growth in the presence of tryptophan. pWT121/FPV 412(L) grew equally well in both sets of conditions. The results of radioimmunoassay of these cultures are shown in Table 2. Because of the differences in growth rate, the values are all normalised to the A_{600} of the culture containing pWT121/FPV 411(R) plus tryptophan. In pWT121/FPV 411(R), HA production is stimulated 65-fold when the *trp* operon is derepressed. In contrast, IAA decreased HA synthesis in pWT121/FPV 412(L). Thus increased transcription from the *trp* promoter is competing with transcription from the presumed promoter near or around the tetracycline gene and reducing expression in the leftward direction.

The effects of repression and induction on HA synthesis are also seen in Fig. 3. As little as 5 μg ml^{-1} tryptophan was sufficient for maximal repression. On the other hand, because of its effect on growth, the induction characteristics of IAA were more complex. Maximal induction occurred at IAA concentrations of 4–8 μg ml^{-1} and HA synthesis was stimulated 120-fold over the repressed level.

Finally, we examined the immunoprecipitable products in cells containing various plasmids after induction or repression of *trp* transcription. At 3 h after addition of either IAA or tryptophan to cultures of such cells, aliquots were pulse-labelled with ^{35}S-methionine and the immunoprecipitable proteins separated by SDS-polyacrylamide gel electrophoresis. No immunoprecipitable proteins were present in induced cells containing pWT121 or in repressed cells containing pWT121/FPV 411(R) (Fig. 4*a* and *b*, respectively). However, induced cells containing pWT121/FPV 411(R) or pWT111/FPV 502(R) and repressed cells containing

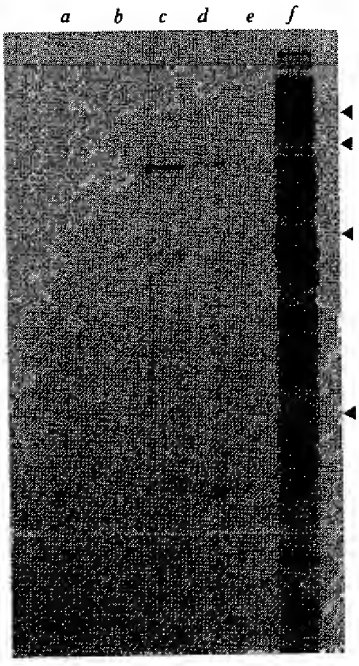


Fig. 4 Immunoprecipitation of the haemagglutinin-like protein. Cultures of *E. coli* K12 HB101 containing the plasmids pWT121, pWT121/FPV 411(R), pWT121/FPV 412(L) and pWT111/FPV 502(R) were grown in M9 salts, glucose, casamino acids medium containing carbenicillin (100 μg ml^{-1}) and either IAA (10 μg ml^{-1}) or tryptophan (40 μg ml^{-1}). After a 3-h period, 5-ml samples were removed and pulse-labelled for 10 min at 37 °C with 50 μCi of ^{35}S-methionine. The labelling was terminated by the addition of 20 ml cold M9 medium, the cells pelleted by centrifugation at 12,000*g* for 10 min at 4 °C and lysed, in a final volume of 1 ml, as described in Fig. 2*A*. Immunoprecipitations were performed as follows: to 400 μl *E. coli* extract was added 2 μl 10% Nonidet P-40 (NP-40), 4 μl (2 μg) normal rabbit IgG and the mixture incubated for 60 min at 20 °C. An aliquot of 10 μl of a 10% suspension of killed *Staphylococcus aureus*[31] (Cowan 1 strain) was then added and incubation continued for a further 15 min. The resulting immunoprecipitate was pelleted and discarded. To the supernatant was added 2 μg rabbit anti FPV-HA and, after 60 min at 20 °C, 10 μl of *S. aureus*. This immunoprecipitate was pelleted 15 min later and washed twice with a solution containing 0.15 M NaCl, 5 mM EDTA, 50 mM Tris-HCl *p*H 7.4 and 0.05% NP-40. Radioactive proteins were eluted from the complex with 30 μl SDS buffer, heated 90 °C for 2 min and electrophoresed on a 12.5% polyacrylamide-SDS gel[32] which was then dried and autoradiographed. The figure shows the immunoprecipitable proteins from cells containing the following plasmids: *a*, pWT121 + IAA; *b*, pWT121/FPV 411(R) + *trp*; *c*, pWT121/FPV 411(R) + IAA; *d*, pWT111/FPV 502(R) + IAA; *e*, pWT121/FPV 412(L) + *trp*; *f* contains an extract from induced cells containing pWT121/FPV 411(R). The positions of conalbumin (MW 76,000), bovine serum albumin (69,000), ovalbumin (43,000) and carbonic anhydrase (30,000) are indicated by the arrows.

pWT121/FPV 412(L) all produced a band of MW 61,000 (Fig. 4*c*, *d* and *e* respectively). Our experimental design predicted that the HA-like protein synthesised from pWT121/FPV 411(R) would have a MW of 69,168 and that the products from pWT121/FPV 502(R) and pWT121/FPV 412(R) would be 26 amino acids smaller at the N-terminus. As the three products are in fact the same size, it seems probable that the primary gene product has been processed and a portion of the polypeptide removed. The same results were obtained when the products of these plasmids were examined in minicells (unpublished results). Clearly the mechanism of the processing and the possible involvement of the pre-HA sequence need further investigation.

Conclusions

We have demonstrated the feasibility of producing controlled amounts of influenza antigenic determinants by genetic engineering. Obviously further analysis is necessary to characterise the protein product and a number of questions remain unanswered. First, for example, initiation of protein synthesis from the prokaryotic-like ribosome binding site on the HA will result in a polypeptide having a eukaryotic leader sequence. This sequence is thought to function in the transmembrane movement of the HA *in vivo*[26]; we may question whether this sequence is recognised by the regulatory systems for export in bacteria. Second, what effect do the $(T_{19}):(A_{19})$ region, resulting from the oligo(dT) primer and the poly(A)tail of the vRNA and situated between the AUG of *trpE* and the HA sequence, or the $(Phe)_6$ oligopeptide resulting from its translation, have on transcription and translation respectively? Homopolymeric nucleotide regions such as this are one of the signals implicated in transcription termination[27]; thus the $(T_{19}):(A_{19})$ region may have a polar effect on HA production. Similarly, the $(Phe)_6$ region may result in depletion of charged $tRNA_{Phe}$ and slow ribosome movement along the mRNA. Clearly examination of HA production by plasmids where this region has been removed is essential.

The absolute quantity of HA-like protein produced may also be estimated. The estimates based on immunological detection cannot be precise and are probably minimum values since if there are differences from the natural antigen it seems likely that the bacterial product would have fewer antigenic determinants or lower affinity for the antiserum. Based on our maximum yield of 60 ng per 50 μl bacterial lysate (Table 2) and a protein concentration of 160 μg ml^{-1} for the same lysate we arrive at a figure of 0.75% of total protein. By comparison, the quantities of induced proteins can be determined by densitometry from autoradiographs of gels containing proteins from induced and repressed cells. This shows that the 61,000 MW protein is 2–3% of total protein synthesis (unpublished results). Neither of these values was as high as expected. Hallewell and Emtage[28] have recently cloned an *E. coli Hin*dIII fragment containing the *trp* promoter and the complete *trpE* gene in pBR322; in induced conditions this plasmid was capable of specifying up to 30% of total protein as anthranilate synthetase[28]. Our values however are similar to those reported for the synthesis of ovalbumin-like proteins under the control of the *lac* promoter[29,30]; that is, the level of expression of the eukaryotic genes is only about 10% of the expected. Whether this is due to limiting amounts of some tRNA species as proposed by Fraser and Bruce[30] or to the type of transcription termination discussed above remains to be seen.

We thank Drs J. Oxford and M. Gethin for gifts of rabbit anti-FPV HA and *Staphylococcus aureus* respectively and Dr A. Hale for providing research facilities. Cloning and expression experiments were performed in a CIII laboratory as recommended by GMAG.

Received 1 October; accepted 22 November 1979.

1. Schild, G. C. in *Chemoprophylaxis and Virus Infections of the Respiratory Tract* Vol. 1 (ed. Oxford, J.S.) 63–101 (CRC Press, Cleveland, 1977).
2. Kavet, J. thesis, Harvard Univ. (1972).
3. Drzeniek, R., Seto, J. T. & Rott, R. *Biochim. biophys. Acta* **128**, 547–558 (1966).
4. Laver, W. G. & Kilbourne, E. D. *Virology* **30**, 493–501 (1966).
5. Wiley, D. C., Skehel, J. J. & Waterfield, M. D. *Virology* **79**, 446–448 (1977).
6. Porter, A. G. *Nature* **282**, 471–477 (1979).
7. Skehel, J. J. & Schild, G. C. *Virology* **44**, 396–408 (1971).
8. Laver, W. G. *Virology* **45**, 275–288 (1971).
9. Tacon, W. A., Carey, N. H. & Emtage, J. S. *Molec. gen. Genet.* (in the press).
10. Lee, F., Bertrand, K., Bennett, G. & Yanofsky, C. *J. molec. Biol.* **121**, 193–217 (1978).
11. Steitz, J. A. & Jakes, K. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4734–4738 (1975).
12. Glover, D. in *New Techniques in Biophysics and Cell Biology* Vol. 8 (eds Pain, R. H. & Smith, B. J.) 125–145 (Wiley, New York, 1976).
13. Miller, J. H. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, New York, 1976).
14. Weissman, C. & Boll, W. *Nature* **261**, 428–429 (1976).
15. Sharp, P. A., Sugden, B. & Sambrook, J. *Biochemistry* **12**, 3055–3063 (1973).
16. Sutcliffe, J. G. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3737–3741 (1978).
17. Broome, S. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2746–2749 (1978).
18. Livingston, D. M. *Meth. Enzym.* **34**, 723–731 (1974).
19. Hunter, W. M. & Greenwood, F. C. *Biochem. J.* **91**, 43–46 (1964).
20. Rodriguez, R. L., West, R. W., Heyneker, H. L., Bolivar, F. & Boyer, H. W. *Nucleic Acids Res.* **6**, 3267–3287 (1979).
21. Widera, G., Gautier, F., Lindenmaier, W. & Collins, J. *Molec. gen. Genet.* **163**, 301–305 (1978).
22. Boyer, H. W. *et al.* in *Recombinant Molecules: Impact on Science and Society* (eds Beers, R. F. & Bassett, E. G.) 9–20 (Raven, New York, 1971).
23. Genetic Engineering: New Guidelines for U.K. *Nature* **276**, 104–108 (1978).
24. Platt, T. in *The Operon* (eds Miller,, J. H. & Reznikoff, W. S.) 263–302 (Cold Spring Harbor Laboratory, New York, 1978).
25. Squires, C. L., Lee, F. D. & Yanofsky, C. *J. molec. Biol.* **92**, 93–111 (1975).
26. Waterfield, M. D., Espelie, K., Elder, K. & Skehel, J. J. *Br. med. Bull.* **35**, 57–63 (1979).
27. Adhya, S. & Gottesman, M. *A. Rev. Biochem.* **47**, 967–996 (1978).
28. Hallewell, R. A. & Emtage, J. S. *Gene* (in the press).
29. Mercereau-Puijalon, O. *et al. Nature* **275**, 505–510 (1978).
30. Fraser, T. H., & Bruce, B. J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5936–5940 (1978).
31. Kessler, S. W. *J. Immun.* **115**, 1617–1624 (1975).
32. Laemmli, U. K. *Nature* **227**, 680–685 (1970).

TAB AA

Proc. Natl. Acad. Sci. USA
Vol. 76, No. 2, pp. 760–764, February 1979
Biochemistry

A general method for maximizing the expression of a cloned gene

(exonuclease III/nuclease S1/*lac* promoter/deletion formation/*cro* protein)

THOMAS M. ROBERTS, RAYMOND KACICH, AND MARK PTASHNE

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138

Communicated by Paul Doty, December 6, 1978

ABSTRACT We present a method, utilizing a combination of restriction endonuclease cleavage and digestion with *Escherichia coli* exonuclease III and *Aspergillus orizae* nuclease S1, that allows us to position a restriction fragment bearing the promoter of the *lacZ* gene of *E. coli* at virtually any distance in front of any cloned gene. In particular, we have used this method to examine the effect on protein production of gene–promoter separation for the *cro* gene of phage λ and to produce plasmids that, upon transformation into appropriate *E. coli* hosts, direct the synthesis of up to 190,000 *cro* protein monomers per cell.

Previous reports from this laboratory have described the use of a DNA fragment generated by restriction endonuclease cleavage as a "portable promoter." This fragment bears the promoter of the *lac* operon of *Escherichia coli* and is capable of directing efficient transcription *in vivo* of genes to which it has been fused *in vitro*. The fusions are carried on plasmids derived from pMB9 (1, 2). In the case of the *cI* gene of phage λ, the amount of *cI* protein (λ repressor) produced by a strain bearing a given *cI–lac* promoter fusion was a sensitive function of the gene–promoter separation. Strains bearing one of these plasmids, pKB280, direct the synthesis of over 30,000 monomers of *cI* protein per transformed cell, or roughly 150-fold more repressor than is found in a typical λ lysogen. In pKB280 the DNA fragment containing the *lac* promoter was abutted to the end of the *cI* gene so that the ribosome binding site (see *Discussion*) of the fusion was a hybrid of *lac* and λ sequences. Fusions which placed the *lac* promoter at a considerably greater distance from the *cI* gene produced 1/5th to 1/10th the amount of *cI* protein. In these experiments, however, the actual number of fusions explored was limited by the availability of convenient restriction cuts in close proximity to the 5′ terminus of *cI*.

In this communication, we describe a method that, in principle, will allow the same *lac* promoter fragment to be placed at virtually any distance in front of a gene. The promoter fragment does not encode a translational start. It does, however, encode a sequence required for binding the message to the ribosome (see *Discussion*). Therefore, our gene–promoter fusions will produce a native protein rather than a fusion protein carrying foreign amino-terminal amino acids (3, 4). In particular, we present details of the construction of a series of *lac* promoter fusions to the *cro* gene of λ designed to examine systematically the effect on protein production of gene–promoter separation, and to produce a strain that synthesizes large amounts of *cro* protein. The best of these strains directs the synthesis of 190,000 monomers of *cro* protein per cell.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

EXPERIMENTAL PLAN

Our basic approach is to clone a gene in a plasmid such that a unique restriction endonuclease cleavage site is located near (within approximately 100 base pairs of) the 5′ end of the gene. We then open the plasmid at that site and excise varying amounts of DNA with exonuclease III and the single-strand-specific endonuclease S1. We then insert a small DNA fragment bearing the promoter of the *lac* operon of *E. coli* and close the plasmid. This produces a set of plasmids bearing the promoter separated by varying distances from the gene. The specific procedure we used to construct the various *lac* promoter–*cro* gene fusions is diagrammed in Fig. 1. As indicated in the next paragraph, many of these steps might easily be modified to accommodate differing configurations of restriction enzyme cuts surrounding different genes. The essential features of the experiment of Figure 1 are as follows: (*i*) the *cro* gene was cloned: a DNA restriction fragment 550 base pairs long and bearing the *cro* gene was sheared to roughly 260 base pairs. Chemically synthesized *Bam* linkers (5) were ligated to the ends of this shorter fragment, and the resulting fragment was cloned into the *Bam* site of pBR322 (6) yielding pTR116. (*ii*) The *Bam* cut at the carboxy end of the *cro* gene in pTR116 was removed: the plasmid was partially digested with *Bam*; the resulting *Bam* sticky ends were rendered flush with *E. coli* DNA polymerase I and the four deoxynucleotide triphosphates (1). This left a single *Bam* cut in the resulting plasmid (pTR151) 54 bases from the amino terminus of the *cro* gene. (*iii*) Some or all of the 54 base pairs of DNA between the *Bam* cut and the ATG signaling the start-point of translation were removed: pTR151 was cut with *Bam* and resected for various times with *E. coli* exonuclease III with subsequent treatment with *Aspergillus orizae* nuclease S1 to remove single-stranded tails. (*iv*) The plasmid was cut at the unique *R* I site some 375 base pairs upstream from the *Bam* site, and an *R* I–*Alu* restriction fragment containing the *lac* promoter was inserted into the plasmid backbone. After transformation into *E. coli*, the resulting plasmids were characterized as described below.

In theory, the *cro* gene in the above construction could have been any gene, and the *Bam* site could have been any unique restriction site. If the *Bam* site had been located inconveniently far from the beginning of the gene, it could have been moved closer by opening the plasmid with *Bam*, digesting with *Exo* III and S1, and then religating the resulting plasmid in the presence of an excess of *Bam* linkers. The *R* I site utilized in the *cro* construction can also be substituted for by several other restriction sites on pBR322 (e.g., *Pst*, *Bam*, *Hin* III, or *Sal* I). We wish to emphasize the convenience of the construction used: the most difficult step, the original cloning of the gene, is done once and then left unchanged. The only inserted fragment, the *lac* fragment, bears a *lac* operator. Cells bearing plasmids

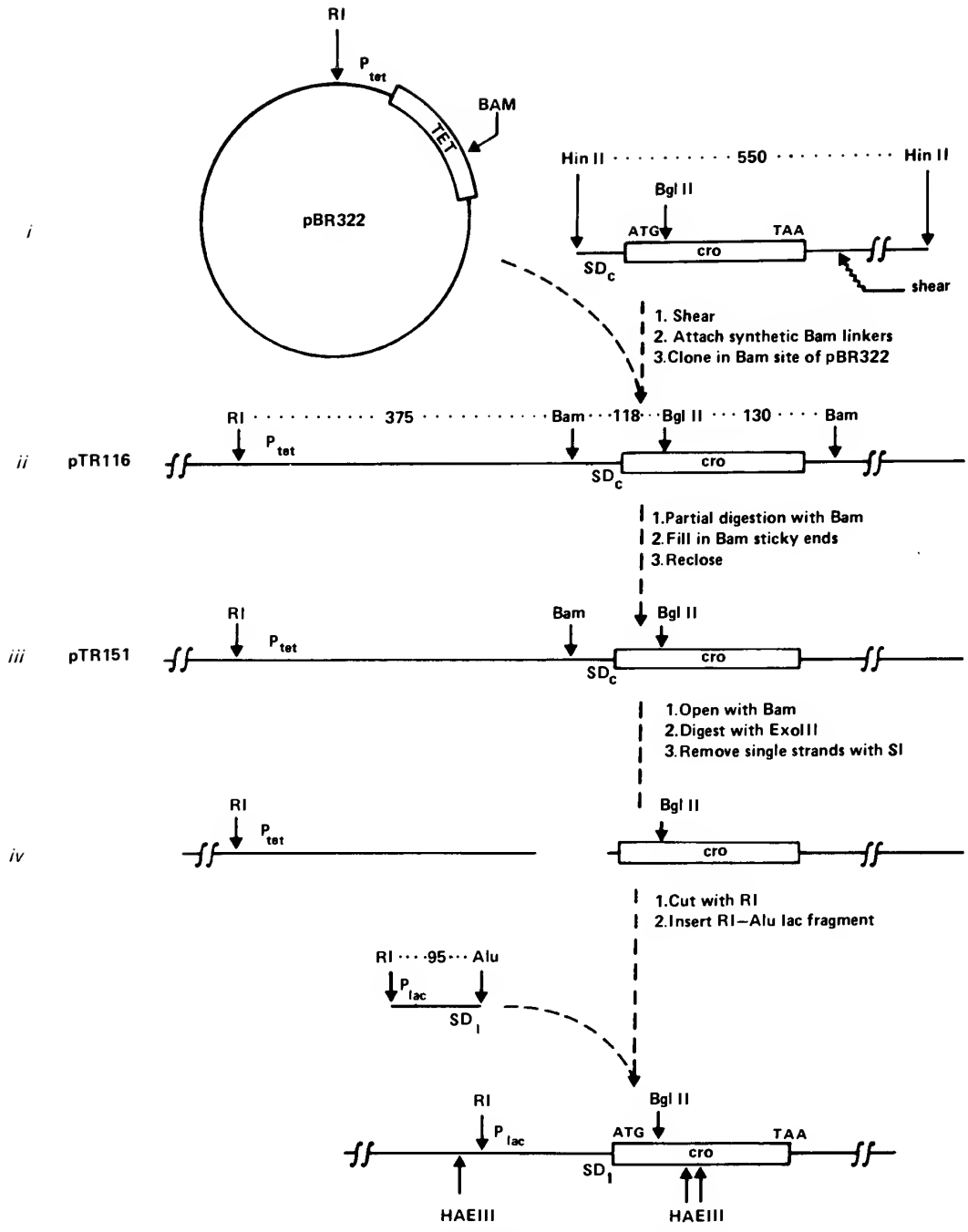


FIG. 1. Schematic representation of the methods of plasmid construction. The approximate locations of several restriction endonuclease cleavage sites are shown for the plasmid pBR322, for a DNA fragment hearing the *cro* gene of phage λ, and for a DNA fragment bearing the promoter of the *lacZ* gene [see Backman and Ptashne (2) for the source of this fragment]. The location of the *tet* and *lacZ* gene promoters are indicated, as are the extent of the *tet* and *cro* genes. SD_c and SD_l indicate the Shine–Dalgarno sequences of the *cro* and *lacZ* genes, respectively. AUG and UAA are the start and stop signals for translation of the *cro* protein. Distances are indicated in hase pairs. Steps: (*i*) The fragment bearing the *cro* gene was shortened hy shearing to remove certain λ control elements near the 3′ end of the gene, and the smaller fragment was inserted into the *Bam* site in pBR322 by using *Bam* linkers. (*ii*) The *Bam* site near the carboxy terminus of the *cro* gene was eliminated. (*iii*) The plasmid was opened at the *Bam* site and varying amounts of DNA were removed hy *Exo* III and S1. (*iv*) The partially resected plasmid was cut at the *R* I site, and the *lac* promoter (bearing the UV5 mutation rendering it independent of catabolite activator) was inserted by "sticky end" ligation at its *R* I end and by "blunt end" ligation to the resected plasmid DNA at its *Alu* end. The efficiency of steps *iii* and *iv* are fairly high—about 200–400 plasmids result from 1 μg of pTR151 input.

carrying the operator are easily recognized on indicator plates. Plasmids bearing the *lac* operator cause colonies of cells to turn blue on the appropriate indicator plate, because multiple copies of the operator titrate the *lac* repressor resulting in synthesis of β-galactosidase (see ref. 1).

MATERIALS AND METHODS

Strains. *E. coli* strain 294 (*Endo I*, $r_k^-m_k^+B_1^-pro^-$) was used as the host for all plasmids. Cells were grown in TB medium supplemented with 0.1% yeast extract (7). Phage strains λKH54 (8), λ4*V'S* (V_2 V_{305} V_{3c} V_{S236}) (9), and λc411 [selected from 4*V'S* by growth on strains producing large amounts of *cro* (J. Eliason, unpublished results)] were used for cross-streaking.

Enzymes. T4 ligase and *E. coli* DNA polymerase I were used as described in Backman *et al.* (1). *Bgl* II restrictions were carried out in 6.6 mM Tris·HCl, pH 7.4/6.6 mM $MgCl_2$/6.6 mM 2-mercaptoethanol. *Bam*, *R I*, and *Hae* III digestions were carried out in the same buffer plus 60 mM NaCl. Exonuclease III (10) digestions were carried out at 22.5°C in the same buffer used for *Bam* digestion in 25–50-μl reaction mixtures at a DNA concentration of 100 μg/ml by using eight units of exonuclease III per μg of DNA. Under these conditions *Exo* III digests at a rate of 8–10 base pairs per min per end depending on the batch of enzyme. Exonuclease III digestion was halted by addition of an equal volume of two times concentrated S1 buffer: 100 mM NaOAc–HOAC, pH 4.0/300 mM NaCl/12 mM $ZnSO_4$. S1 nuclease was added to a concentration of 75 units per 50 μl and allowed to digest for 2 hr at 18°C (11). The reaction was stopped by phenol extraction, and the DNA was purified over a G-50 fine Sephadex column before further reaction.

Plasmid Constructions. *pTR116.* The 550-base-pair *Hin* II restriction fragment (5 μg) bearing the *cro* gene was dialyzed against 100 mM NaOAc, pH 8.0/1 mM EDTA/66.7% glycerol in a final volume of 20 ml. The DNA was sheared for 30 min at 40,000 rpm in a VirTis homogenizer. A dry ice/isopropanol bath was used for cooling. After ethanol precipitation of the DNA, polymerase I was used to fill in the single-stranded regions generated by shearing (providing the single-stranded projection had a 5′ end). *Bam* linkers (5) were then ligated onto ends of the shear fragments (75 pmol of linkers in a total volume of 20 μl). After digestion with *Bam*, the fragments (now bearing *Bam* sticky ends) were cloned into the *Bam* site of pBR322. Transformed cells were selected for ampicillin resistance and λ immunity (1). DNA was purified (12) from a small number of candidates, and the sequence of the inserted *Bam* piece from the plasmid with the smallest insert (pTR116) was determined (13). The insert extended from the *Hin* II site at the left end of the original restriction fragment to a point five bases past the 3′ end of the *cro* gene (data not shown). The shearing procedure thus neatly separated the *cro* gene from certain λ control elements near the carboxy terminus of the gene that we wished to exclude from our future constructions.

pTR151. pTR116 DNA (5 μg) was partially digested with *Bam*, and those DNA molecules receiving only one *Bam* cut were separated from the other species present by agarose gel electrophoresis. After extraction from the gel (14), this DNA was treated with DNA polymerase 1 and the four deoxynucleotide triphosphates to render flush the *Bam* sticky ends (1). The resulting molecules were circularized with T4 ligase and used to transform *E. coli*. DNA was isolated from several ampicillin-resistant, λ-immune colonies, and restriction analysis was carried out to determine which of the product plasmids had lost the desired *Bam* site (data not shown).

pTR161. pTR151 DNA (5 μg) was opened with *Bam*. DNA polymerase 1 was used to fill in the resulting sticky ends. The DNA was then cut with *R* 1 and the *R* I–*Alu* promoter fragment (in 3-fold molar excess) was ligated into place. All plasmids were constructed under conditions conforming to the standards outlined in the National Institutes of Health guidelines.

Radioimmunoassay. *cro* protein was measured in lysates of transformed cells by radioimmunoassay using the procedure of A. Johnson (unpublished results). Radioimmunoassay measurements were quite precise—there was, on the average, less than 5% difference between duplicate measurements on a culture. The amount of *cro* protein was found to be a function of cell growth conditions. However, measurements made on separate occasions on cultures of independent transformants bearing a given plasmid differed by less than 35%, and the ratio of *cro* protein present in strains bearing two different plasmids varied less than 10% from day to day. Gel electrophoresis of total soluble protein from strains bearing plasmid pTR214 gave estimates of *cro* protein as a percent soluble protein in accord with the radioimmunoassay determination (data not shown) —this indicates that the majority of the *cro* protein detected by radioimmunoassay is intact.

DNA Sequence Analysis. DNA from each of the various λ–*lac* fusion plasmids was digested with *Bgl* II, which cleaves at a unique site 67 bases downstream from the amino terminus of the *cro* gene. The resulting linear DNA was 3′-end labeled by using DNA polymerase I and [α-^{32}P] dGTP and [α-^{32}P] dATP (15). Subsequent *R* I digestion released a small DNA fragment, 162–220 base pairs long, bearing the region of fusion of *lac* and λ sequences. This fragment was isolated by polyacrylamide gel electrophoresis, and the sequence through the fusion region was determined by the method of Maxam and Gilbert (13).

RESULTS

pTR151 DNA was opened with *Bam*, resected with exonuclease III for times ranging from 30 sec to five min, and then treated with S1 nuclease (16). An *R* I–*Alu* fragment bearing the *lac* promoter was ligated into place. After transformation, cell cultures were plated on ampicillin selective plates containing the indicator 5-chloro-4-bromo-3-indolyl-β-D-galactoside. Blue colonies, arising from cells that bear the *lac* promoter on the transforming plasmid and are thus *lac* constitutive, were then checked for *cro* protein production by cross-streaking against phage λ. A typical experiment yielded 400 colonies per μg of pTR151 DNA, 70% of which are blue. Of these blue colonies, 10–90% were immune. Our assumption, later confirmed by radioimmunoassay, was that increasingly high levels of *cro* protein production would confer resistance to ever more virulent derivatives of λ. Clones were tested for immunity to phages bearing no operator mutations (λKH54) or four or more operator mutations (λ4*V'S*, λc411). In this way, the transformants were divided into categories by level of immunity. DNA was isolated from some 40 transformants, representative of the various classes, and characterized by digestion with *Hae* III (see Fig. 2). Because the entire sequence of pBR322 is known (17) as well as the sequences of the *lac* promoter fragment (18) and the *cro* gene itself (19), the *Hae* digests were readily interpreted. The size of the deletion made by exonuclease III and S1 nuclease treatment was estimated from the size of the fragment containing the *lac–cro* fusion.

Finally, nine representative transformants were chosen for more detailed analysis. The level of *cro* protein produced by each of these clones was measured by radioimmunoassay, and the DNA sequence across the *lac–cro* fusion was determined. Fig. 3 shows the results of this analysis.

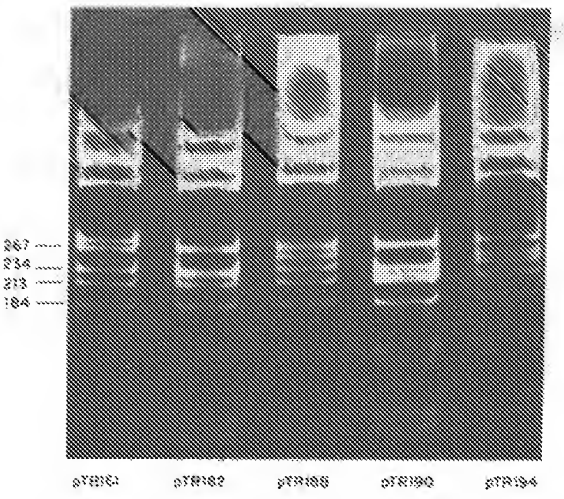


FIG. 2. Analysis of *Hae* III digests of plasmid DNA by gel electrophoresis. Plasmid DNAs from the various *cro* protein-producing strains were partially purified and digested with *Hae* III. Digests were electrophoretically separated on 8% polyacrylamide gels. Of the *Hae* III fragments produced from any plasmid, all but one fragment are common to all plasmids. The one varying fragment contains the region of fusion between *lac* and λ sequences. It can be sized by comparison to the other fragments, the whole sequences of which are all known. The lengths in base pairs of four of these pBR322 fragments are indicated.

DISCUSSION

Strains bearing two of our fusions, pTR213 and pTR214, are rather spectacular overproducers of *cro* protein. The levels of *cro* protein as a percent of total soluble protein shown in Fig. 3 of 1.6% and 1.0% correspond to roughly 190,000 and 120,000 monomers of *cro* per cell. We know of no other *E. coli* strains which produce any protein to a significantly greater extent on a molar basis. If there are 50 copies of pTR213 per cell [the copy number of pBR322 (6)], each *lac* promoter must be directing the production of 4000 monomers of *cro*. This is roughly the fully induced level of expression of β-galactosidase from the same promoter when it directs transcription of the *lac* operon (2). We have not excluded the possibility that these fusions have acquired additional mutations that change the number of copies of plasmid maintained in the cell. The high frequency with which fusions are produced that direct the synthesis of large amounts of protein argues strongly against this possibility, however.

The *lac–cro* fusions presented here represent the beginning of a systematic examination of the effect of gene–promoter separation on protein production. The nine plasmids described in Fig. 3 all carry a complete *cro* gene and a complete *lac* promoter, separated by varying distances. However, the variation in the amount of *cro* protein produced by strains bearing the different fusion plasmids is enormous. For example, strains carrying pTR213 produce over 2000-fold more *cro* protein than strains carrying pTR190. Current theory offers plausible explanations for some of the observed differences, but fails to explain all of them. The same promoter is being used in each plasmid, and so we assume that transcription across the *cro* gene is uniform in each case, and therefore the differences in protein productions are due to differences in some post-transcriptional process. The *cro* mRNAs transcribed from the various plasmids differ in their leader regions, and these differences might affect mRNA stability or processing or ribosome binding efficiency. We do not know how the presence of different leaders might influence the stability or processing of a message. However, current theory holds that a ribosome binding site on a prokaryotic mRNA is composed of two parts: the AUG or GUG signaling the start-point of translation and the so-called Shine and Dalgarno sequence—3–9 bases in the leader portion of the message that are complementary to bases at the 3′ end of the 16S rRNA (21, 22). In six of the nine plasmids, the bases coding for the *cro* Shine–Dalgarno sequence are intact (pTR161, 213, 199, 214, 188, and 194), and it seems reasonable to assume that this sequence is functioning in ribosome binding. Although relatively high levels of *cro* protein are produced in cells bearing all of these plasmids, there are striking differences among them. For example, the deletion in pTR199 is only three bases longer than that in pTR213 and five bases shorter than that in pTR214; yet, strains transformed with pTR199 direct synthesis of 1/10th of the protein of either of the others. In pTR210 one of the bases of the *cro* Shine–Dalgarno sequence has been removed, and yet strains carrying this plasmid produce 10-fold higher levels of *cro* protein than strains carrying pTR199. We have no explanation for these differences. In the two remaining plasmids, the bases coding for the *cro* Shine–Dalgarno sequence are either

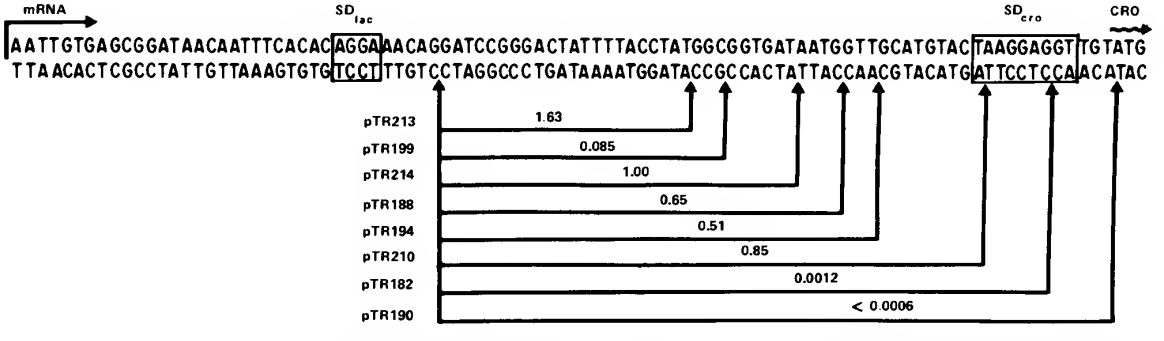


FIG. 3. A summary of the deletion size and *cro* production of selected plasmids. Shown is a portion of the sequence of pTR161 extending from the start-point of transcription of the *lac* promoter on the left (straight arrow) to the start-point of translation of the *cro* gene on the right (wavy arrow). All plasmids produced by the *Exo* III–S1 method described in Fig. 1 may be considered deletions of pTR161. The extent of these deletions is indicated by brackets. The numbers on the brackets are the level of *cro* protein as a percentage of total soluble protein in cells transformed with each of the deletion plasmids. For comparison, *cro* protein represents 0.5% of the total soluble protein in cells transformed with pTR161, and a previously reported plasmid (20) in which *cro* transcription is initiated at its own promoter directs synthesis of 0.05% *cro*. By cross-streaking, strains bearing pTR182 and pTR190 are immune to λKH54; and strains bearing pTR213, 214, 188, 194, and 210 are immune to λKH54, λ4*V*′S, and to λc411, a λ strain selected from λ4*V*′S for growth on *E. coli* strains producing large amounts of *cro*.

largely (pTR182) or totally (pTR190) absent. Cells transformed with these plasmids produce extremely small amounts of *cro* protein. This is perhaps understandable in light of the absence of the *cro* Shine–Dalgarno sequences; however, it must be remembered that presumptive ribosome binding sites in pTR182 and pTR190 are actually coded for by a fusion of *lac* and λ sequences. The *lac* promotor fragment used in the construction of all the plasmids described here contains, in addition to the binding site for RNA polymerase and the start-point of transcription, the coding region for most of the leader of the *lacZ* message including its Shine–Dalgarno sequence. Thus, in pTR182 and pTR190, the *lac* Shine–Dalgarno sequence is brought into close proximity with the ATG of *cro*. We know that such hybrid ribosome binding sites can be extremely effective from the results with pKB280 (2). It is possible that the separation of this sequence and the ATG in pTR190 and pTR182 is not optimum—the *lac* sequence is separated from the ATG of *cro* by 10 bases in pTR182 and by 5 bases in pTR190, whereas the same sequence is separated from the ATG of the *cI* gene by 8 bases in pKB280 and is separated from the ATG of the *lacZ* gene by 7 bases in the *E. coli* chromosome. Work is in progress to produce a *lac–cro* fusion in which the *lac* Shine–Dalgarno sequence is seven bases from the ATG of *cro*.

The method of exonuclease III and S1 nuclease digestion used here should allow the placement of the promoter-containing fragment at virtually any distance upstream from most other genes. As we mentioned earlier, the requirements for restriction sites in and around other genes of interest are extremely flexible. Perhaps the only limitation in promoter placement suggested by the sequence data in Fig. 3 is a tendency of the deletion to stop at As or Gs in preference to Ts in the 5′ (top) strand (there are not enough Cs available to make a judgment). In particular, the technique should be useful in positioning the *lac* promoter fragment adjacent to a eukaryotic gene with the hope of forming a hybrid ribosome binding site from the *lac* Shine–Dalgarno sequence and the ATG of the eukaryotic gene. Expression of the eukaryotic gene could be monitored by radioimmune techniques (23, 24). It remains to be seen whether such fusions will produce eukaryotic proteins in *E. coli*.

We thank A. Johnson for purified *cro* protein and anti-*cro* serum, J. Eliason for phage stocks, A. Jeffrey for restriction enzymes *Hae* III and *R* I, W. McClure for *E. coli* exonuclease III and polymerase I, and R. Scheller for *Bam* linkers.

1. Backman, K., Ptashne, M. & Gilbert, W. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 4174–4178.
2. Backman, K. & Ptashne, M. (1978) *Cell* **13**, 65–71.
3. Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3727–3731.
4. Mercereau-Puijalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, F. & Kourilsky, P. (1978) *Nature (London)* **275**, 505–510.
5. Scheller, R. H., Dickerson, R. E., Boyer, H. W., Riggs, A. D. & Itakura, K. (1977) *Science* **196**, 177–180.
6. Boyer, H. W., Betlach, M., Bolivar, F., Rodriguez, R. L., Heyneker, H. L., Shine, J. & Goodman, H. M. (1977) in *Recombinant Molecules: Impact on Science and Society*, eds. Beers, R. F., Jr. & Basset, E. G. (Raven, New York), pp. 9–20.
7. Miller, J. H., ed. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
8. Blattner, F. R., Fiandt, M., Hank, K., Twore, P. A. & Szybalski, W. (1974) *Virology* **62**, 458–471.
9. Flashman, S. (1976) Dissertation (Harvard University, Cambridge, MA).
10. Richardson, C. C., Lehman, I. R. & Kornberg, A. (1964) *J. Biol. Chem.* **239**, 251–258.
11. Wu, R., Ruben, G., Siegel, B., Jay, E., Spielman, P. & Tu, C. D. (1976) *Biochemistry* **15**, 734–740.
12. Clewell, D. B. & Helinski, D. R. (1970) *Biochemistry* **9**, 4428–4440.
13. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
14. Thuring, R. W. S., Sanders, J. P. M. & Borst, P. (1975) *Anal. Biochem.* **66**, 213–221.
15. Schwartz, E., Scherer, G., Hobom, G. & Kossel, H. (1978) *Nature (London)* **272**, 410–414.
16. Carbon, J., Shenk, T. E. & Berg, P. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1392–1396.
17. Sutcliffe, J. G. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43**, in press.
18. Gilbert, W., Gralla, J., Majors, J. & Maxam, A. (1975) in *Protein-Ligand Interactions*, eds. Sund, H. & Blauer, G. (Walter de Gruyter, Berlin), pp. 193–206.
19. Roberts, T. M., Shimatake, H., Brady, C. & Rosenberg, M. (1977) *Nature (London)* **270**, 274–275.
20. Johnson, A., Meyer, B. J. & Ptashne, M. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1783–1787.
21. Shine, J. & Dalgarno, L. (1975) *Nature (London)* **254**, 34–38.
22. Steitz, J. A. (1978) in *Biological Regulation and Development*, ed. Goldberger, R. (Plenum, New York), Vol. 1, pp. 349–399.
23. Erlich, H. A., Cohen, S. N. & McDevitt, H. O. (1978) *Cell* **13**, 681–689.
24. Broome, S. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2746–2749.

TAB BB

United States Patent [19]

Ptashne et al.

[11] **4,332,892**

[45] **Jun. 1, 1982**

[54] **PROTEIN SYNTHESIS**

[75] Inventors: **Mark Ptashne; Gail D. Lauer; Thomas M. Roberts,** all of Cambridge, Mass.; **Keith C. Backman,** San Francisco, Calif.

[73] Assignee: **President and Fellows of Harvard College,** Cambridge, Mass.

[21] Appl. No.: **111,101**

[22] Filed: **Jan. 10, 1980**

Related U.S. Application Data

[63] Continuation of Ser. No. 3,102, Jan. 15, 1979, abandoned.

[51] **Int. Cl.**[3] **C12P 21/00**

[52] **U.S. Cl.** **435/68;** 435/70; 435/172

[58] **Field of Search** 435/172, 70, 71, 68

[56] **References Cited**

PUBLICATIONS

Chang et al., Nature, vol. 275, pp. 617–624, Oct. 19, 1978.

Itakura et al., Science, vol. 198, pp. 1056–1063 (1977).

Backman et al., Proc. Natl. Acad. Sci., USA, vol. 73, pp. 4174–4178 (1976).

Backman et al., Cell, vol. 13, pp. 65–71, Jan. 1978.

Primary Examiner—Alvin E. Tanenholtz

[57] **ABSTRACT**

This invention is a process to produce specific proteins coded for by eukaryotic (or prokaryotic) DNA in bacteria. The invention, which uses recombinant DNA techniques, produces proteins in their natural, functional state unencumbered by extraneous peptides.

6 Claims, No Drawings

PROTEIN SYNTHESIS

This is a continuation of application Ser. No. 3,102, filed Jan. 15, 1979, and now abandoned.

The invention described herein was made in the course of work under a grant or award from the National Science Foundation.

This invention is a method for producing in bacteria prokaryotic or eukaryotic proteins in native, unfused form free from extraneous peptides.

Recombinant DNA techniques in vitro have been used to insert a variety of eukaryotic genes into plasmids carried by *Escherichia coli* in an effort to induce these bacteria to produce eukaryotic proteins. Most of these genes have not directed the synthesis of the native proteins because the eukaryotic signals coding for initiation of transcription and/or translation do not function well in *E. coli*. One proposed solution to this problem has been the fusion of the eukaryotic gene with a bacterial gene. The process results in the production of a hybrid protein, a portion of which at its carboxyl terminus is constituted by the eukaryotic protcin. In one case, it has been possible to separate a small biologically active protein from a fusion product (Itakura, K. et al., *Science* 198, 1056 (1977)).

Gene expression takes place by transcription into mRNA then translation into protein. To do these operations, the DNA preceding the gene must have a sequence which: (a) directs efficient binding of bacterial RNA polymerase and efficient initiation of transcription, and (b) codes for a mRNA that directs efficient binding of mRNA to the ribosomes and initiation of translation into protein.

The present invention provides a method of producing native, unfused prokaryotic or eukaryotic protein in bacteria which comprises inserting into a bacterial plasmid a gene for a prokaryotic or eukaryotic protein and a portable promoter consisting of a DNA fragment containing a transcription initiation site recognized by RNA polymerase and containing no protein translational start site, said promoter being inserted ahead of a protein translational start site of said gene to form a fused gene having a hybrid ribosome binding site, inserting said plasmid into said bacteria to transform said bacteria with said plasmid containing said fused gene, and culturing the transformed bacteria to produce said prokaryotic or eukaryotic protein.

The present invention utilizes nucleases, restriction enzymes, and DNA ligase to position a portable promoter consisting of a DNA fragment containing a transcription site but no translation initiation site near the beginning of the gene which codes for the desired protein to form a hybrid ribosomal binding site. The protein produced by the bacterium from this hybrid is the native derivative of the implanted gene. It has been found that the endonuclease digestion product of the *E. coli lac* operon, a fragment of DNA which contains a transcription initiation site but no translational start site, has the required properties to function as a portable promoter in the present invention, being transcribed at high efficiency by bacterial RNA polymerase. The mRNA produced contains a Shine-Dalgarno (S-D) Sequence but it does not include the AUG or GUG required for translational initiation. However, in accordance with the present invention, a hybrid ribosomal binding site is formed consisting of the S-D sequence and initiator from the lac operon and the ATG sequence of the gene, and such a fused gene is translated and transcribed efficiently. Using the enzymes exonuclease III and S1, the promoter may be put at any desired position in front of the translational start site of the gene in order to obtain optimum production of protein. Since the promoter can be inserted at a restriction site ahead of the translational start site of the gene, the gene can first bc cut at the restriction site, the desired number of base pairs and any single stranded tails can be removed by treating with nucleases for the appropriate time period, and religating.

The following specific example is intended to illustrate more fully the nature of thc present invention without acting as a limitation upon its scope.

EXAMPLE

A rabbit β-globin gene was first cloned into the Hin III site of pBR322, a plasmid of the *E. coli* bacteria, via restriction enzyme cuts of the initial DNA, reconstitution of the gene by T4 ligase, insertion of the reconstituted gene into the Hin III site using chemically synthesized Hin III linkers, and religating with DNA ligase.

The Hin III cut at the carboxyl end of the cloned gene was removed by partially digesting with Hin III, filling in the resulting Hin III "sticky ends" with *E. coli* DNA polymerase I, and religating with T4 ligase. This left in the resulting plasmid a single Hin III cut 25 base pairs ahead of the amino terminus of the globin gene.

Differing number of the 25 base pairs between the Hin III cut and the ATG signalling the start point of translation were removed from different samples of the cloned gene as follows: the plasmid was cut with Hin III, resected for various times from 0.5 to 10 minutes with Exo III, then treated with S1 to remove single-stranded tails.

The portable promoter of the lac operon, an R1-Alu restriction fragment of *E. coli* DNA, was then inserted by treating each sample of the plasmid with R1 which cuts at a unique site some 30 base pairs upstream from the Hin III site, and the portable promoter was insertec into the plasmid backbone at this site. This requires one "sticky end" and one "flush" end, both of which are ligated by the same treatment with ligase.

Colonies of *E. coli* each containing one of these resulting plasmids were then screened for β-globin production using RIA-screening techniques to identify the one or more producing β-globin.

The globin gene in the above construction can be any gene coding for prokaryotic or eukaryotic proteins, and any other unique restriction site can be employed in place of the Hin III site. If the restriction site is located inconveniently far from the beginning of the gene, it may be moved (for example, a Hin III site may be moved by opening the plasmid with Hin III, digesting with Exo III and S1, then religating the resulting plasmid in the presence of excess Hin III linkers). Any suitable restriction site can be employed for insertion of the portable promoter in place of the R1 site (e.g. Pst, BAM, or Sal I). Finally, it should be emphasized that the most difficult step, the cloning of the gene into the plasmid, is done once and left unchanged. The promoter fragment will confer its constitutive expression on the cell so it is easy to screen for the intact promoters.

What is claimed is:

1. The method of producing native, unfused prokaryotic or eukaryotic protein in bacteria which comprises inserting into a bacterial plasmid a gene for a prokaryotic or eukaryotic protein and a portable promoter

consisting of a DNA fragment containing a Shine-Dalgarno sequence and a transcription initiation site recognized by RNA polymerase and containing no protein translational start site, said promoter being inserted upstream from a protein ATG translational start site of said gene, at a position to obtain production of said protein, to form a fused gene having said Shine-Dalgarno sequence, said transcription initiation site, and the ATG signalling the start point of translation, inserting said plasmid into said bacteria to transform said bacteria with said plasmid containing said fused gene, and culturing the transformed bacteria to produce said unfused prokaryotic or eukaryotic protein.

2. The method as claimed in claim **1** in which said bacteria is *E. coli*.

3. The method as claimed in claims **1** or **2** in which said portable promoter is the product of restriction endonuclease digestion of an operon.

4. The method claimed in claim **2** in which said portable promoter is the product of restriction endonuclease digestion of the lac operon of *E. coli*.

5. The method as claimed in claims **1** or **2** in which said gene insertion comprises cloning said gene into said plasmid, adjusting the spacing between said inserted gene and a preceding unique restriction site by treating with a nuclease, and cloning said portable promoter into said restriction site.

6. The method as claimed in claim **2** in which said gene insertion comprises cloning said gene into said plasmid, adjusting the spacing between said gene and a preceding unique restriction site by treating with a nuclease, and cloning into said restriction site a portable promoter formed by the endonuclease digestion of the lac operon of *E. coli*.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 4,332,892
DATED : June 1, 1982
INVENTOR(S) : Mark Ptashne et al.

It is certified that error appears in the above-identified patent and that said Letters Patent are hereby corrected as shown below:

Under "Inventors", delete the following names:

"Gail D. Lauer; Thomas M. Roberts"

Signed and Sealed this

Fourteenth Day of *September 1982*

[SEAL]

Attest:

GERALD J. MOSSINGHOFF

Attesting Officer *Commissioner of Patents and Trademarks*

TAB CC

United States Patent [19]

Ptashne et al.

[11] **4,418,149**

[45] **Nov. 29, 1983**

[54] **FUSED HYBRID GENE**

[75] Inventors: **Mark Ptashne; Gail D. Lauer; Thomas M. Roberts,** all of Cambridge, Mass.; **Keith C. Backman,** San Francisco, Calif.

[73] Assignee: **President and Fellows of Harvard College,** Cambridge, Mass.

[21] Appl. No.: **346,084**

[22] Filed: **Feb. 5, 1982**

Related U.S. Application Data

[60] Division of Ser. No. 111,101, Jan. 10, 1980, Pat. No. 4,332,892, which is a continuation of Ser. No. 3,102, Jan. 15, 1979, abandoned.

[51] **Int. Cl.**[3] **C12N 1/20;** C12N 15/00; C12P 21/00; C07H 21/04

[52] **U.S. Cl.** **435/253;** 435/68; 435/172; 536/27

[58] **Field of Search** 435/172, 68, 317, 253; 536/27

[56] **References Cited**

PUBLICATIONS

Dickson et al., Science vol. 187, pp. 27–35 (1975).

Metzler, Biochemistry, The Chemical Reactions of Living Cells, Academie Press Inc., pp. 907–910 (1977).

Backman et al., Cell, vol. 13, pp. 65–71, Jan. 1978.

Chang et al., Nature, vol. 275, pp. 617–624, Oct. 19, 1978.

Itakura et al., Science, vol. 198, pp. 1056–1063 (1977).

Backman et al., Proc. Natl. Acad. Sci. USA, vol. 73, pp. 4174–4178 (1976).

Primary Examiner—Alvin E. Tanenholtz

[57] **ABSTRACT**

This invention is a process to produce specific proteins coded for by eukaryotic (or prokaryotic) DNA in bacteria. The invention, which uses recombinant DNA techniques, produces proteins in their natural, functional state unencumbered by extraneous peptides.

8 Claims, No Drawings

FUSED HYBRID GENE

This is a division, of application Ser. No. 111,101, filed Jan. 10, 1980, now U.S. Pat. Not. 4,332,892, which is a continuation of Ser. No. 3,102 filed Jan. 15, 1979 now abandoned.

The invention described herein was made in the course of work under a grant or award from the National Science Foundation.

This invention is a method for producing in bacteria prokaryoctic or eukaryotic proteins in native, unfused form free from extraneous peptides.

Recombinant DNA techniques in vitro have been used to insert a variety of cukaryotic genes into plasmids carried by *Escherichia coli* in an effort to induce these bacteria to produce eukaryotic proteins. Most of these genes have not directed the synthesis of the native proteins because the eukaryotic signals coding for initiation of transcription and/or translation do not function well in *E. coli.* One proposed solution to this problem has been the fusion of the eukaryotic gene with a bacterial gene. The process results in the production of a hybrid protein, a portion of which at its carboxyl terminus is constituted by the eukaryotic protein. In one case, it has been possible to separate a small biologically active protein from a fusion product (Itakura, K. et al., *Science* 198, 1956 (1977)).

Gene expression takes place by transcription into mRNA then translation into protein. To do these operations, the DNA preceding the gene must have a sequence which: (a) directs efficient binding of bacterial RNA polymerase and efficient initiation of transcription, and (b) codes for a mRNA that directs efficient binding of mRNA to the ribosomes and initiation of translation into protein.

The present invention provides a method of producing native, unfused prokaryotic or eukaryotic protein in bacteria which comprises inserting into a bacterial plasmid a gene for a prokaryotic or eukaryotic protein and a portable promoter consisting of a DNA fragment containing a transcription initiation site recognized by RNA polymerase and containing no protein translational start site, said promoter being inserted ahead of a protein translational start site of said gene to form a fused gene having a hybrid ribosome binding site, inserting said plasmid into said bacteria to transform said bacteria with said plasmid containing said fused gene, and culturing the transformed bacteria to produce said prokaryotic or eukaryotic protein.

The present invention utilizes nucleases, restriction enzymes, and DNA ligase to position a portable promoter consisting of a DNA fragment containing a transcription site but no translation initiation site near the beginning of the gene which codes for the desired protein to form a hybrid ribosomal binding site. The protein produced by the bacterium from this hybrid is the native derivative of the implanted gene. It has been found that the endonuclease digestion product of the *E. coli* lac operon, a fragment of DNA which contains a transcription initiation site but no translational start site, has the required properties to function as a portable promoter in the present invention, being transcribed at high efficiency by bacterial RNA polymerase. The mRNA produced contains a (Shine-Dalgarno (S-D) Sequence) but it does not include the AUG or CUG required for translational initiation. However, in accordance with the present invention, a hybrid ribosomol binding site is formed consisting of the S-D sequence and initiator from the lac operon and the ATG sequence of the gene, and such a fused gene is translated and transcribed efficiently. Using the enzymes exonuclease III and S1, the promoter may be put at any desired position in front of the translational start site of the gene in order to obtain optimum production of protein. Since the promoter can be inserted at a restriction site ahead of the translational start site of the gene, the gene can first be cut at the restriction site, the desired number of base pairs and any single standard tails can be removed by treating with nucleases for the appropriate time period, and religating.

The following specific example is intended to illustrate more fully the nature of the present invention without acting as a limitation upon its scope.

EXAMPLE

A rabbit β-globin gene was first cloned into the Hin III site of pBR322, a plasmid of the *E. coli* bacteria, via restriction enzyme cuts of the initial DNA, reconstitution of the gene by T4 ligase, insertion of the reconstituted gene into the Hin III site using chemically synthesized Hin III linkers, and religating with DNA ligase.

The Hin III cut at the carboxyl end of the cloned gene was removed by partially digesting with Hin III, filling in the resulting Hin III "sticky ends" with *E. coli* DNA polymerase I, and religating with T4 ligase. This left in the resulting plasmid a single Hin III cut 25 base pairs ahead of the amino terminus of the globin gene.

Differing numbers of the 25 base pairs between the Hin III cut and the ATG signalling the start point of translation were removed from different samples of the cloned gene as follows: the plasmid was cut with Hin III, resected for various times from 0.5 to 10 minutes with Exo III, then treated with S1 to remove single-standard tails.

The portable promoter of the lac operon, an R1-Alu restriction fragment of *E. coli* DNA, was then inserted by treating each sample of the plasmid with R1 which cuts at a unique site some 30 base pairs upstream from the Hin III site, and the portable promoter was inserted into the plasmid backbone at this site. This requires one "sticky end" and one "flush" end, both of which are ligated by the same treatment with ligase.

Colonies of *E. coli* each containing one of these resulting plasmids were then screened for β-globin production using RIA-screening techniques to identify the one or more producing β-globin.

The globin gene in the above construction can be any gene coding for prokaryotic or eukaryotic proteins, and any other unique restriction site can be employed in place of the Hin III site. If the restriction site is located inconveniently far from the beginning of the gene, it may be moved (for example, a Hin III site may be moved by opening the plasmid with Hin III, digesting with Exo III and S1, then religating the resulting plasmid in the presence of excess Hin III linkers). Any suitable restriction site can be employed for insertion of the portable promoter in place of the R1 site (e.g. Pst, BAM, or Sal I). Finally, it should be emphasized that the most difficult step, the cloning of the gene into the plasmid, is done once and left unchanged. The promoter fragment will confer its constitutive expression on the cell so it is easy to screen for the intact promoters.

What is claimed is:

1. A fused hybrid gene capable of expressing native unfused prokaryotic or eukaryotic protein consisting

essentially of (1) a portable promoter including a portion of a bacterial gene having a Shine-Dalgarno sequence and a transcription initiation site recognized by RNA polymerase and containing no protein translational start site, and (2) fused thereto a gene for a native unfused prokaryotic or eukaryotic protein including its translational start site, said portable promoter being located upstream from said translational start site.

2. A fused gene as claimed in claim **1** in which said promoter is the product of restriction endonuclease digestion of an operon.

3. A fused gene as claimed in claim **2** in which said bacterial gene is that of *E. coli* and said promoter is the product of restriction endonuclease digestion of the lac operon of *E. coli*.

4. A fused gene as claimed in claim **1** wherein said translational start site of said gene for said prokaryotic or eukaryotic protein is the sequence ATG.

5. A fused gene as claimed in claim **1** wherein said portable promoter and said translational start site of said prokaryotic or eukaryotic gene together comprise a hybrid ribosomal binding site.

6. A fused gene as claimed in claim **1** wherein said gene for a prokaryotic or eukaryotic protein is a gene for a eukaryotic protein.

7. A bacterium containing the fused gene of claim **1**, said bacterium being capable of producing said prokaryotic or said eukaryotic protein.

8. A bacterium containing the fused gene of claim **6**, said bacterium being capable of producing said eukaryotic protein.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 4,418,149

DATED : November 29, 1983

INVENTOR(S) : Mark Ptashne et al.

It is certified that error appears in the above-identified patent and that said Letters Patent are hereby corrected as shown below:

Col. 1, line 28, "1956" should be --1056--;

Col. 1, line 66, "CUG" should be --GUG--;

Col. 2, line 37, "standard" should be --stranded--.

Signed and Sealed this

Seventh Day of *February 1984*

[SEAL]

Attest:

GERALD J. MOSSINGHOFF

Attesting Officer *Commissioner of Patents and Trademarks*